

-199-

5	orf75-1	ATLADMAELFPERRIMLAREITKTFETFLSGTVGEIQTALSADGNQSRGEVVLVLYPAQD	190	200	210	220	230	240
	orf75a.pap	EKHEGLSESAQINIMKILTAEPLTKQAELAAKITGEGKKALYDLALSWKNKX	250	260	270	280	290	
	orf75-1	EKHEGLSESAQINIMKILTAEPLTKQAELAAKITGEGKKALYDLALSWKNKX	250	260	270	280	290	
10 Homology with a predicted ORF from <i>N.gonorrhoeae</i>								
ORF75 shows 93.2% identity over a 292aa overlap with a predicted ORF (ORF75.ng) from <i>N.gonorrhoeae</i> :								
15	orf75.pap	MFVFQTAFFMFOKHLQKASDVSVGGTLYVATPIGNLADITLRLALVLQKA---AEDTR	56					
	orf75.ng	MSVFQTAFFMFOKHLQKASDVSVGGTLYVATPIGNLADITLRLALVLQKADICAEEDTR	60					
20	orf75.pap	VPAQLLSAYGIQGLVSVREHNERQMAKIVGYLSDGMVVAQVSADGTPAVCDPGAKLAR	116					
	orf75.ng	VTAQLLSAYGIQGLVSVREHNERQMAKIVGYLSDGLVVAQVSADGTPAVCDPGAKLAR	120					
25	orf75.pap	RVREAGFKVVPVVGAKAVMAALSVAEGSDFYFNGFVPPKSGERRKLFKXWVRAAFPIV	176					
	orf75.ng	RVREAGFKVVPVVGAKAVMAALSVAEGSDFYFNGFVPPKSGERRKLFKXWVRAAFPIV	180					
30	orf75.pap	MFETPHRIGAAALADMAELFPERRIMLAREITKTFETFLSGTVGEIQTALSADGNQSRGEM	236					
	orf75.ng	MFETPHRIGATLADMAELFPERRIMLAREITKTFETFLSGTVGEIQTALAADGNQSRGEM	240					
35	orf75.pap	VLVLVPAQDEKHEGLSESAQINIMKILTAEPLTKQAELAAKITGEGKKALYD	288					
	orf75.ng	VLVLVPAQDEKHEGLSESAQINIMKILAAELPTKQAELAAKITGEGKKALYDLALSWKNK	300					

An ORF75ng nucleotide sequence <SEQ ID 291> was predicted to encode a protein having amino acid sequence <SEQ ID 292>:

35	1	MSVFQTAFFM FOKHLQKASD SVVGGTLYVV ATPIGNLADI TRLALVLQK
	51	ADIICAEEDTR VTAQLLSAYG IQGRLVSVRE HNERQMAKIV IGFLSDGLV
40	101	AQVSDAGTTPA VCDPGAKLAR RVREAGFKVPV PVVGASAVMA ALSVAGVAES
	151	DFYFNGFVPP KSGERRKLEFA KXWVRAAFPIV MFETPHRIGA TLADMAELFP
45	201	ERRIMLAREI TKTFTETFLSG TVGEIQTALA ADGNQSRGEM VLVLVPAQDE
	251	KHEGLSESAQ NAMKILAAEL PTKQAELAA KITGEGKKAL YDLALSWKNK
	301	*

After further analysis, the following gonococcal DNA sequence <SEQ ID 293> was identified:

45	1	ATGTTTCAGA AACACTTGCA GAAAGCCTCC GACACGCTCG TCGGAGGGAC
	51	ATTATACGCTG GTTGCCACGC CCAATCGGCA TTTGGCGAGC ATTACCTGCG
50	101	CGCCTTTGGC GGTATTGCAA AAGGCGGACA TCATTGTGCG CGAAGACAGC
	151	CGCCTTACTG CCGAGCTTTT GAGCGGCTAC CGCAITCAGG CAGAGTTGGT
55	201	CAGTGTGGCG GAACACACCG AGCGCGCAGT GCGCGACAAG GATATCGGTT
	251	TCCTTTTCAGA CGCCTGTGTT GTGCGCGAGG TTTCCGATCG GGTACGCCCG
60	301	CGCGTGTGCG ACCCGGGCGC GAAATCTGCC CGCCGCGTGC GCGAAGCAGG
	351	GPTCAAAAGTC GTTCCGCTCG TGGCGCGAAG CGCGGTAAAG CGCGCGTTGA
65	401	GTGTGGCCGG GTGTGGCGGA TCCGATTTTT ATTTCAACGG TTTTGTACCG
	451	CCGAAATCGG GCGAAGCTAG GAAATTTGTT GCCAAATGGG TCGGGCGGCG
70	501	ATTTCCTGTC GTCAATGTTT AAGCGCGGCA CGGAATCGGG GCAACGTTTG
	551	CCGATATGGC GGAATGTTTC CCGCAACGCC GTCTGATGCT GCGCGCGGAA
75	601	ATCACGAAAA CGTTTGAACG GTTCTTAAGC GGCACGGTTG GGGAAATTCA
	651	GACGGCATTC GCGCGGGAGC GCAACCAATC GCGCGGGAGC ATGGTGTGTC
80	701	TGCTTTTATCC GCGCGAGGAT GAAAAACAGC AAGGCTGTGC CGAGTCTGCG
	751	CAAAATGCGA TGAATAATCT TCGCGCGGAG CTGCGCGACA AGCAGCGGCG
85	801	GGAGCTTGCC GCGAAGATTA CAGTGCAGGG CAAAAAGGCT TTGTACGATT
	851	TGGCACTGTC GTGGAAAAAC AATATA

60 This corresponds to the amino acid sequence <SEQ ID 294; ORF75ng-1>:

-200-

	1	MFQKHLQKAS	DSVVGGLTYV	VATPIGNLAD	ITLRALAVLQ	KADIIICAEDT	
	51	RVTAQLLSAY	GIQGRILSVR	EHNERQMAK	VIGFLSDGLV	VAQVSDAGT	P
	101	AVCDPGAKLA	RRVREAGFKV	VPVVGASAVM	AALSVAGVAE	SDFYFNGFVP	
	151	PKSGERRKLF	AKWVRAAFV	VMFETPHRIG	ATLADMAELF	PERRLMLARE	
5	201	ITKTETFTLS	GTVGEIQTAL	AADGNQSRGE	MVLVLYPAQD	EKHEGLSESA	
	251	QNAKILAAE	LPTKQAELA	AKITGEGKKA	LYDLALSWKN	K*	
ORF75ng-1 and ORF75-1 show 96.2% identity in 291 aa overlap:							
		10	20	30	40	50	60
10	orf75-1.pep	MFQKHLQKASDSVVGGLTYV	VATPIGNLADITLRALAVLQ	KADIIICAEDTRVTAQLLSAY			
	orf75ng-1	MFQKHLQKASDSVVGGLTYV	VATPIGNLADITLRALAVLQ	KADIIICAEDTRVTAQLLSAY			
		10	20	30	40	50	60
		70	80	90	100	110	120
15	orf75-1.pep	GIQGKLVSVREHNERQMAK	IVGYLSDGMVVAQVSDAGT	PAVCDPGAKLARRVREAGFKV			
	orf75ng-1	GIQGKLVSVREHNERQMAK	IVGYLSDGMVVAQVSDAGT	PAVCDPGAKLARRVREAGFKV			
		70	80	90	100	110	120
		130	140	150	160	170	180
20	orf75-1.pep	VPVVGASAVMAALSVAGV	ESDFYFNGFVPPKSGERRKLF	AKWVRAAFVVMFETPHRIG			
	orf75ng-1	VPVVGASAVMAALSVAGV	ESDFYFNGFVPPKSGERRKLF	AKWVRAAFVVMFETPHRIG			
		130	140	150	160	170	180
25		190	200	210	220	230	240
	orf75-1.pep	ATLADMAELFPERRLMLARE	ITKTETFTLSGTVGEIQTAL	SADGNQSRGEMVLVLYPAQD			
	orf75ng-1	ATLADMAELFPERRLMLARE	ITKTETFTLSGTVGEIQTAL	SADGNQSRGEMVLVLYPAQD			
		190	200	210	220	230	240
30		250	260	270	280	290	
	orf75-1.pep	EKHEGLSESAQNMKILTAEL	PTKQAELAIAKITGEGKALY	LDLALSWKNKX			
35	orf75ng-1	EKHEGLSESAQNMKILTAEL	PTKQAELAIAKITGEGKALY	LDLALSWKNKX			
		250	260	270	280	290	

Furthermore, ORF75ng-1 shows significant homology to a hypothetical *E. coli* protein:

	sp P45528 YRAL_ECOLI HYPOTHETICAL 31.3 KD PROTEIN IN AGAI-MTR INTERGENIC REGION (F286)	
40	>q1606086 (U18997) ORF_f286 [Escherichia coli]	
	>q1789535 (AE000395) hypothetical 31.3 kD protein in agai-mtr intergenic region [Escherichia coli] length = 286	
	Score = 218 bits (550), Expect = 3e-56	
	Identities = 128/284 (45%), Positives = 171/284 (60%), Gaps = 4/284 (1%)	
45	Query: 4 KHLQKASDSVVGGLTYVATPIGNLADITLRALAVLQKADIIICAEDTRVTAQLLSAYGIQ 63	
	K Q A + S G LY-V TPIGNLADIT RAL VLO D-I AEDTR T LL +GI	
	Sbjct: 2 KQHSADNSQ--GQLYIVPTPIGNLADITQRALEVLQAVDLIAEDTRHTGLLQHPGIN 59	
50	Query: 64 GRLVSVREHNERQMAKIVGYLSDGLVVAQVSDAGTPAVCDPGAKLARRVREAGFKVVPV 123	
	RL ++ +HNE+Q A ++ L +G -A VSDAGT + DPG L R REAG +VVP+	
	Sbjct: 60 ARLFALHDHNEQQKAETLLAKLQEGQNTALVSDAGTPLINDPGYHLVTRCERAGIRVVP 119	
55	Query: 124 VGASAVMAALSVAGVAESDFYFNGFVPPKSGERRKLFKAWVRAAFVVMFETPHRIGATL 183	
	G A + ALS AG+ F +GF+P KS RR ++ +E+ HR+ +L	
	Sbjct: 120 PGPCAAITALSAGLPSDRFCYEGFLPAKSKGRDRDALKAIIEAPRTLI FYESTHRLLDLSL 179	
60	Query: 184 ADMAELFPERR-LMLAREITKTETFTLSGTVGEIQTALAADGNQSRGEMVLVLYPAQDE 242	
	D+ + E R ++LARE+TKT+ET VGE+ + D N+ +GEMVL++ +	
	Sbjct: 180 EDIVAVLGESESVYVLAELTKTWETIHGAFVGEALLAVKEDENRRKGEMVLIV-EGHKAQ 238	
	Query: 243 HEGLSAQNMKILTAELPTKQAELAIAKITGEGKALYDLAL 286	
	E L A + +L AELP +AA LAA+I G K ALY AL	
65	Sbjct: 239 EEDLPADALRTIALQAEPLKAAALAAIEHGVKKNALYKYAL 282	

Based on this analysis, including the presence of a putative transmembrane domain in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 35

- 5 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 295>:

```

1 ATGAAACAGA AAAAAACCGC TGCOCGAGTT ATTGCTGCAA TGTTCGGCAGG
51 TTTTGCGGCA GC.AAAGCAC CGAATATCGA CCGGCTTTG .....
//
651 ..... ..GAGTTGG TCAGAAACCA GTTGAGCAGG GGTTCAGAC
10 701 AGGAAAAAGC CGCCTTGAAA ATCGATGCC TTTTGAAGA AAACGGTGTG
751 AAACCGTAA

```

This corresponds to the amino acid sequence <SEQ ID 296; ORF76>:

```

1 MKQKKTAAAV IAAMLAGFAA XKAPEIDPAL .....
15 201 ..... ELVRNQLQQ LRQEKARLKI DALLEENGVK
251 P*

```

Further work revealed the complete nucleotide sequence <SEQ ID 297>:

```

1 ATGAAACAGA AAAAAACCGC TGCOCGAGTT ATTGCTGCAA TGTTCGGCAGG
51 TTTTGCGGCA GCCAAAGCAC CGAATATCGA CCGGCTTTG GTGGATAOCG
20 101 TGGTGGCGCA GATCATGCAG CAGGCAGACC GGCATCGCGA GCAGTCCCAA
151 AAAACCGGACG GGCAGGCAAT CGAATATCGA CCGGCTCGCC GCCTACAAAC
201 TTTGGAAGTT TTGAAAAACA GGGCATTGAA GGAAGGTTTG GATAAGGATA
251 AGGATGTCCA AAACCGCTTT AAATTCGCCG AAGCGTCTTT TTATGCCGAG
301 GAGTAAGTCC GTTTTCTGGA ACGTTCCGAA ACGGTTTCOC AAGACGAGCT
25 351 GCACAAAGTTT TAGCAACAGC AAATCCGCTAT GATCAAAATTG CAGCAGGTCA
401 GCTTCGCAAC CGAAGAGGAG CGCGCTCAGG CCGCAGCTCT CTGCTCAAA
451 GAGCTGTCTT TTGAAGGGCT GATGAAGCCT TATCCGAAGC ACGACAGGC
501 TTTTGAAGCT TTCAATTATGG CGCAGCAGCT TCCCGAGCGC CTGCTTCGCG
551 AGTTTCGCCG GATGAATCGG GGCAGCTTAA CCGCGATGCT GGTCAAAATTG
30 601 GCGCAAGCCT ATTATCTGTT CAAACTCAGC GAGGTGCGGA AAAACCCCGA
651 CGCGCAGCCT TTCAGATTGG TCAGAAACCA GTTGAGCAGG GGTTCAGAC
701 AGGAAAAAGC CGCCTTGAAA ATCGATGCC TTTTGAAGA AAACGGTGTG
751 AAACCGTAA

```

This corresponds to the amino acid sequence <SEQ ID 298; ORF76-1>:

```

35 1 MKQKKTAAAV IAAMLAGFAA XKAPEIDPAL VDTLVAIQM QADRHAEOSQ
51 KPDGAIRND AVRRLOTLEV LKNRLKEGL DKDKDVQNRF KIAEASFYAE
101 EYVRFERSE TVSEDELHFK YEQOIRMIKL QOVSFATEEE ARQAQQLLLK
151 GLSFEGLMKR YPNDEQAFDG FIMAQQLPEP LASQFAMNR GDVTRDFVCL
40 201 GERYLFLKL EVGKMPDAQP FELVRNQLQ GLRQEKARLK IDALLEENGV
251 KP*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF76 shows 96.7% identity over a 30aa overlap and 96.8% identity over a 31aa overlap with an ORF (ORF76a) from strain A of *N. meningitidis*:

```

45          10      20      30
orf76.pep  MKQKKTAAAVIAAMLAGFAAXKAPEIDPAL
           TTTTTTTTTTTTTTTTTTTTTTTTTT
orf76a     MKQKKTAAAVIAAMLAGFAAXKAPEIDPALVDTLVAIQMQADRHAEOSQKPDGQAIRND
           10      20      30      40      50      60
           //
                    70      80      90

```

```

orf76.pep                                KELVRNQLSEQGLRQEKARLKIDALLEENGVKPX
                                     |||
orf76a      DVTRDPVKLGERYYLFLKSEVGKNPDAQPFELVRNQLSEQGLRQEKARLKIDALLEENGVKPX
               200       210       220       230       240       250

```

5 The complete length ORF76a nucleotide sequence <SEQ ID 299> is:

```

1  ATGAAACAGA  AAAAAACCGC  TGCCGCGAGT  ATTGCTGCAA  TGTGTGGCAGG
51  TTTTGGCGCA  GCCAAAGCAC  CCGAAATCGA  CCGGCTTTG  GTGGATACGC
101 TGGTGGCGCA  GATCATGCGA  CAGGCGAGCC  GGCATGCGGA  GCAGTCCCAA
151 AAACCGGACG  GGCAGGCAAT  CGGAAAGCAT  GCGTCCGTC  GCGTGCMAAC
201 TTTGAGAGTT  TGAAAAAGAA  CGGATGGA   GAAGGTTTG  GATAGGATA
251 AGGATGTC   AAACCGCTTT  AAATGCGCG  AGGCTCTTT  TTATGCGGAG
301 GAGTACGTCC  GTTTCTGGA  ACGTTGCGAA  ACGGTTCCG  AAAGCGCACT
351 GCGTCAGTT  TATGAGCGGC  AAATCGCAT  GATCAAATG  CAGCAGGTCA
401 GCTTCGGAAC  CGAAGAGGAG  GCGCGTCAGG  CGCAGCAGCT  CCGTCTCAA
451 GGGCTGTC   TTGAGGGCT  GATGAAGCT  TATCCGAGC  ACGAGCAGGC
501 TTTTGAOGGT  TTCATTATGG  CGCAGCAGCT  TCCCGAGCG  CTGGCTTCG
551 AGTTTGACGC  GATGAATCG  GCGCAGCTTA  CCGCGATCC  GGTCAAATTG
601 GCGGAAGCGT  ATTATCTGTT  CAACTCAGC  GAGGTCGGA  AAAACCCGCA
651 CGCGACGCT  TTGAGTTGG  TCAGAAACCA  GTTGGAACA  GGTGTCAGAC
701 AGGAAAAAGC  CCGCTTGAAA  ATCGATGCCA  TTTTGAAGA  AAACGCTGTC
751 AAACCGTAA

```

This encodes a protein having amino acid sequence <SEQ ID 300>:

```

1  MKQKKTAAAV  IAAMLAGFAA  AKAPEIDPAL  VDTLVAQIMQ  QADRHAEQSQ
51  KPDGQAIRND  AVRRQLTLEV  LKNRLKEGL  DKDKDVQNR  KIAEASFYAE
101 EYVRFLESE  TVSESAIRQF  YERQIRMIKL  QQVSFATEE  ARQAQQLLLK
151 GLSFEGLMKR  YPNDEQAFDG  FIMAQQLPEP  LASQFAAMNR  GDVTRDPVKL
201 GERYYLFLKS  EVGKNPDAQP  FELVRNQLSEQ  GLRQEKARLK  IDALLEENGK
251 KP*

```

ORF76a and ORF76-1 show 97.6% identity in 252 aa overlap:

```

30  orf76a.pep      10      20      30      40      50      60
      MKQKKTAAAVIAAMLAGFAAAKAP EIDPALVDTLVAQIMQADRHAESQKPDGQAIRND
      |||
orf76-1      MKQKKTAAAVIAAMLAGFAAAKAP EIDPALVDTLVAQIMQADRHAESQKPDGQAIRND
               10      20      30      40      50      60

35  orf76a.pep      70      80      90      100     110     120
      AVRRQLTLEV LKNRLKEGLDKDKDVQNRFKIAEASFYAEYVRFLESE TVSESAIRQF
      |||
orf76-1      AVRRQLTLEV LKNRLKEGLDKDKDVQNRFKIAEASFYAEYVRFLESE TVSESAIRQF
               70      80      90      100     110     120

40  orf76a.pep      130     140     150     160     170     180
      YERQIRMIKLQQVSFAT EEEARQAQQLLLKGLSFEGLMKRYPNDEQAFDGFIMAQQLPEP
      |||
orf76-1      YERQIRMIKLQQVSFAT EEEARQAQQLLLKGLSFEGLMKRYPNDEQAFDGFIMAQQLPEP
               130     140     150     160     170     180

45  orf76a.pep      190     200     210     220     230     240
      LASQFAAMNRGDVTRDPVKLGERYYLFLKSEVGKNPDAQPFELVRNQLSEQGLRQEKARLK
      |||
orf76-1      LASQFAAMNRGDVTRDPVKLGERYYLFLKSEVGKNPDAQPFELVRNQLSEQGLRQEKARLK
               190     200     210     220     230     240

50  orf76a.pep      250
      IDALLEENGKPKX
      |||
orf76-1      IDALLEENGKPKX
               250

```

60 Homology with a predicted ORF from *N.gonorrhoeae*

The aligned aa sequences of ORF76 and a predicted ORF (ORF76.ng) from *N. gonorrhoeae* of the N- and C-termini show 96.7 % and 100% identity in 30 and 31 overlap, respectively:

	orf76-pep	MKQKKTAAAVIAAMLGFAAKAPEIDPAL	30
	orf76ng		60
5	orf76-pep	ELVRNQLQGLRQEKARLKIDALLEENGVPK	251
	orf76ng	VTRNPVKLGERYYLFLKLGAVGKNPDAQPFELVRNQLQGLRQEKARLKIDALLEENGVPK	251

The complete length ORF76ng nucleotide sequence <SEQ ID 301> is:

10	1	ATGAACAGCA	AAAGACCCG	TGCGCAGTT	ATTGCTGCMA	TGTTGSCAGS
	51	TTTTGCGGCA	GCCAAAGCAG	CCGAAATCGA	CCCGCTTTTG	GTGGATAAGC
	101	TGGTGGCGCA	GATCATGCGAG	CAGGCAGACC	GGCATCGGGA	GCAGTCCCAA
	151	AGACCGGACG	GGCAGGCAAT	CGSAAACGAT	CGCGTCCGCC	GGCTGCAAAC
	201	TTTTGAGATT	TTGAARACA	GGGCATTGAA	GGAAGGTTTG	GATAAGGATA
	251	AGGATGTCCA	AAACCGCTTT	AAAATCGCGC	AAGCGTCTTT	TTATGCGCAG
15	301	GAGTACGTCC	GTTTCTGGA	ACGTTGCGAA	ACGGTTCGC	AAAGCGCACT
	351	CGGTACGTTT	TATGAGCGGC	AAATCGGCAT	GATCAAAATT	CAGCAGGTCA
	401	GCTTCGCAAC	CGAAGAGGAG	GGCGTCAGG	CGCAGCAGCT	CCTGCTCAAA
	451	GGCGTGTCTT	TTGAAGGGCT	GATGAAGCGT	TATCCGAACG	ACGAGCAGGC
	501	GTTTCAGCGT	TTTATTATG	CGCAGCAGCT	TCCCGACCGC	CTGGCTTgcg
20	551	agtttgcggg	TATGAACCGT	GGGACGCTTA	CCGCAATCC	GGTCAAAATT
	601	GGCGACGCGT	ATTACCTGTT	CAAACTCGGC	GGGTCGGGA	AAACCCCGCA
	651	CGCGAGCGCT	TTGAGGTTTG	TCAGAAACCA	GTTGGAACAA	GGTTTGAGCG
	701	AGGAAAGACG	CCGCTTGAAA	ATCGATGCC	TTTGGAGa	Aaacgggtgtc
	751	AaacCGTAA				

25 This encodes a protein having amino acid sequence <SEQ ID 302>:

	1	MKQKKTAAAV	IAAMLGFAA	AKAPEIDPAL	VDTLVAQIMO	QADRHAEEQSQ
	51	RPDQGAIRND	AVRRLQTLV	LKNRALKGL	DKDKDVQNR	KIAEASFYAE
	101	EYVRFLESE	TVSESALRQF	YERQIRMIKL	QQVSFATEE	ARQAQQLLLK
	151	GLSFEGLMKR	YPNDEQAFDG	FIMAQQLPEP	LASQFAGMNR	GDVTRNPVKL
30	201	GERYYLFLKG	AVGKNPDAQF	FELVRNQLQ	GLRQEKARLK	IDALLEENGVP
	251	K*				

ORF76ng and ORF76-1 show 96.0% identity in 252 aa overlap

		10	20	30	40	50	60
35	orf76-1-pep	MKQKKTAAAVIAAMLGFAA	AKAPEIDPALVDTLVAQIMO	QADRHAEEQSQ	KPDGQAIRND		
	orf76ng						
		10	20	30	40	50	60
40	orf76-1-pep	AVRRLQTLV	LKNRALKGLDKDKDVQNR	KIAEASFYAEYVRFLESE	TVSEDELHGF		
	orf76ng						
		70	80	90	100	110	120
45	orf76-1-pep	YEQQIRMIKL	QQVSFATEE	ARQAQQLLLK	GLSFEGLMKR	YPNDEQAFDG	FIMAQQLPEP
	orf76ng						
		130	140	150	160	170	180
50	orf76-1-pep	LASQFAGMNR	GDVTRNPVKL	GERYYLFLK	SEVGKNPDAQF	FELVRNQLQ	GLRQEKARLK
	orf76ng						
		190	200	210	220	230	240
55	orf76-1-pep	LASQFAGMNR	GDVTRNPVKL	GERYYLFLK	SEVGKNPDAQF	FELVRNQLQ	GLRQEKARLK
	orf76ng						
		250					
60	orf76-1-pep	IDALLEENGVPKX					
	orf76ng						
		250					

Furthermore, ORF76ng shows significant homology to a *B. subtilis* export protein precursor:

```

sp|P24327|PRSA_BACSU PROTEIN EXPORT PROTEIN PRSA PRECURSOR >gi|96227|pir||S15269
33K lipoprotein - Bacillus subtilis >gi|39782 (X57271) 33kDa lipoprotein
[Bacillus subtilis]
5 >gi|2226124|gnl|PID|e325181 (Y14077) 33kDa lipoprotein [Bacillus subtilis]
>gi|2633331|gnl|PID|e1182997 (299109) molecular chaperonin [Bacillus subtilis]
Length = 292
Score = 50.4 bits (118), Expect = 1e-05
Identities = 48/199 (24%), Positives = 82/199 (41%), Gaps = 32/199 (16%)

10 Query: 70 VLKNRAIKEGLEK-----DKDVQNRFKIAEASF-----YAEYVRFLESETVSE 114
VL ++ LEK DK++ N+ K + Y ++Y++ + E +++
Sbjct: 53 VLTQLVQEKVLCKYKVSDEKIDNKLKEYKTQLGDQYTALEKQYGVKDYLEKQVYELLTQ 112

15 Query: 115 SA-----LRQFYERQIRMIKLQOVSFATEEEARQAQQLLLKGLSFGELMKRYFN 163
A +++++E I+ + A ++ A ++ L KG FE L K Y
Sbjct: 113 KAAKNIKVTDADIKYEWGLKGIKIRASHILVADKKTAEEVEKKLKKEGKFDLAKEYST 172

Query: 164 DEQAQFDG-----FIMAQQLPEPLASQFAAMNRGVDTRDPVKLGERYYFLKLESEVGNKPD 218
D A G G F Q+ E + + G+V+ DPKV Y++ K +E D
20 Sbjct: 173 DSSASKGSDLGWFKAEGQMDTFESKAAFKLKTGEVS-DPVKTQYGYHIKKTEERGKYDD 231

Query: 219 QPFELVRNQLCGLRQKEA 237
EL LEQ L A
25 Sbjct: 232 MKKELKSEVLEQKLNENAA 250

```

Based on this analysis, including the presence of a putative leader sequence and a RGD motif in the gonococcal protein, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF76-1 (27.8kDa) was cloned in the pET vector and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 10A shows the results of affinity purification of the His-fusion protein. Purified His-fusion protein was used to immunise mice, whose sera were used for Western blot (Figure 10B), ELISA (positive result), and FACS analysis (Figure 10C). These experiments confirm that ORF76-1 is a surface-exposed protein, and that it is a useful immunogen.

35 Example 36

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 303>:

```

1 ATGAAAAAAT CTTCCTTAC GCTTGTCTG TATTCTGCT TACTTACCGC
51 CAGCGAAATT GCCTATACC TTGGAATTGG GGATTGAAC CTTACCGCGC
101 GCAAAAATTG CGGAACGTT TGCCTGACA TTTGTGATT CTGCGCTGTA
151 TCTGTTTTCG CGTAATAAGG TGACGCGTT GTTGATTGG GTGTTTTTGG
201 CGTTGAGCAT TATTGCCAAC AATGTGCATT ACGCGGATTA TCAGAGCTGG
251 ATGAGC.... ..//
1201 ..... CAACCGGTAT TCGAGCGACT CCGAGCGGCA
1251 ACTGCGTGTG TGCTATACC TCCGATCATG GCCAGTATGT TCGCGAGAGT
1301 ATCTCAATC AAGGACCGGT CGAGCCGCGC AGCTATCTCG TCGCCGTAGT
1351 GTTCTACAGC CCGGATAAGG CCGTGCAACA GGCTGCCAAC CAGGCTTTTG
1401 CGCCTTGCGA GATTGCCCTT CATCAGCAGC TTTCAAGCTT CCGTATTCAC
1451 ACGTTGGGCT ACGATATGCC GGTTCAGGT TGTGCGGAG GCTCGGTAAC
1501 GGGCAACCTG ATTACGGGTG ATGCAGCGAG CTTGAACAT CGCGACGGCA
1551 AGGCGGAATA TGTTTATCCG CAATGA

```

This corresponds to the amino acid sequence <SEQ ID 304; ORF81>:

```
1 MKKSFLTLVL YSSLLTASEI AYPLELGIET LPAAKIATSF ALTFVIAALY
```

-205-

```

51 LFARNKVTRL LIAVFFAFSI IANNVHYADY QSWMT.....
//
401 ...QTVEQL QKTPDGNWLF AYTSDHGQYV RQDIYNQGTV QPDSYLVPLV
451 LYSFDKAVQQ AANQAFAPCE IAFHQQLSTF LIHTIGYDMP VSGCREGSVT
501 GNLITGDAGS LNIIRDGAKEY VYPQ*

```

Further work revealed the complete nucleotide sequence <SEQ ID 305>:

```

1 ATGAAAAAAT CTTTCCTTAC GCTTGTTCGT TATTCGTCTT TACTTACGGC
51 CAGCGAAATTT GCCTATCGCT TTGTATTGGG GATTGAACCC TTACCGGGCG
101 CAAAAATGGC GGAAGCTTTT GCGTGACAT TTGTGATGTC TGCGGTGAT
151 CTGTTTGGCG GTTATAAGGT GAGCGGTGTC TTGATGGGG TGTTTTTTCG
201 GTTCAGCATT ATTGCCAACA ATGTGCAITTA CCGCGTTCAT CRAAGCTCGA
251 TGACGGGCAT CAATTATTGG CTGATGCTGA AAGAGGTTCAC CGAAGTCGGC
301 AGCGCGGGTG CGTCGATGTT GGATAAGTTG TGGCTGCGTC TGTGTGGGG
351 CGTGTGGGAA GTCATGTTGT TTTCAGCCTT TGCGAAGTTC CGCGCTAAGA
401 CGCATTTTTC TGCCGATATA CTGTTTGCTT TCCTAATGTC GATGATTTTC
451 GTGGCTTCGT TCGACAGGAA ACAAGAGCAC GGTATTTCGC CCAACCGCAC
501 ATACAGCCGC ATCAAGGCCA ATTATTTTCT CTTGCGTTAT TTGTGCGGAC
551 GCGTGTTCGC GTATCAGTTG TTTGATTAA GCAGGATTCG CGCCTTTAAG
601 CAGCCTGCTC CAAGCAAAAT CGGCGAGGCG AGTGTTCAAA ATATCGTCTT
651 GATTATGGGC GAAAGCGAAA GCGCGGCGCA TTGAAGCTG TTGGCTACG
701 GACGCGAAAC TTCGCCCTTT TTAACCGCGC TGTCGCAAGC CGATTCTAAG
751 CGGATTTGGA AGCAAGATTG TTCCGAGCGC TTATGACGTC CAGTGTCCCT
801 GCGCATTTT TTCAATGCGA TACCGCAGCG CAACGCGCTG GAACAAATCA
851 CGCGCGGCGA TACCAATATG TTCGCCCTCG CAAAGAGGCA GGCCTATGAA
901 ACGTATTTTT ACAGCGCGCA GCGCGAAAC GAGATGCGGA TTTTGAACCT
951 AATCGGTAAG AATAGGATAG ACCATCTGAT TCAGCGGAGC CAACTTGGCT
1001 ACGCGAACGG CGCAATATG CCGCATGAGA AGCTGCTGCC GTTGTGAC
1051 AAAATCAATT TGACAGCAGG CAAGCATTTT ATCGTGTGTC ACCAAGCGGG
1101 TTTCGACGCG CCATACGGCG CATTGTTGCA GCGTCAAGAT AAGATTTCG
1151 GCGAAGCGCA TATTGTGGAT AAGTACGACA ACACCATCCA CAAAACCGAC
1201 CAAATGATTC AAACCGTATT CGAGCAGCTG CAAAGCAGC CTGACGGCAA
1251 CTGGCTGTTT GCCTATACCT CCGATCATGG CAGTATGTTT GCGCAAGATA
1301 TCTACAATCA AGGACGGGTG CAGCGCGACA GCTATCTGCT CGCGTAGTG
1351 TTGTACAGCC CGGATAGGCC CGTGCACAGC GCTCGCAAGC AGGCTTTGTC
1401 GCTTGGGAG AGCGCTTTC ATCAGCAGCT TGCACAGCT CTGATTCACA
1451 CGTTGGGCTA CGATATGCGG GTTTCAGGTT GTCGCAAGC CTCGGTAACG
1501 GGCACACTGA TTACGGGTGA TGCAGGCGAG TTGAACATTC GCGACGCAA
1551 GCGGAAATAT GTTATCCGC AATGA

```

This corresponds to the amino acid sequence <SEQ ID 306; ORF81-1>:

```

40 1 MKKSFLTLVL YSSLTASEI AYRFVGIET LPAAKIAETF ALTFVIAALY
51 LFARYKVTRL LIAVFFAFSI IANNVHYAVY QSWMTGINYV IMLKEVTEVG
101 SAGASMLDKL WLPVLWGVL VMLFCSLAKF RKRTHFSADI LFAFIMLMIF
151 VRSFDTKQEH GISPKPTYSR IKANYFSFY FVGRVLPYQL FDLSPAPAFK
201 QPAPSKIGQG SVQNVILIMG ESESAHLKL FGYGRTSPF LTRLSQADFK
45 251 FIVKQSYSG FMTAVSLPSF FNAI PHANGL EQISGGDTNM FLRAKEQGYE
301 TFEYSAQAEH EMAILNLIGK KWIDHLQPT QLG'YNGNDNM PDEKLLPLFD
351 KIMLQQRHF IYLVQRGSHA FYGALLQPD KYVSEADYD KYDNTFKTD
401 QNLQVTEQL QKTPDGNWLF AYTSDHGQYV RQDIYNQGTV QPDSYLVPLV
451 LYSFDKAVQQ AANQAFAPCE IAFHQQLSTF LIHTIGYDMP VSGCREGSVT
50 501 GNLITGDAGS LNIIRDGAKEY VYPQ*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF81 shows 84.7% identity over a 85aa overlap and 99.2% identity over a 121aa overlap with an ORF (ORF81a) from strain A of *N. meningitidis*:

```

55      10      20      30      40      50      60
orf81.pep MKKSFLTLVL YSSLTASEIAYR FVGIETLPAAKIAETFALTFVIAAL LFARNKVTRL
      |||:::| | | | | | | | | | | | | | | | | | | | | | | | | | | |
orf81a MKKSFLVFLV YSSLTASEIAYR FVGIETLPAAKIAETFALTFVIAAL YLFARYKATRL
      10      20      30      40      50      60
              70      80

```

-206-

5	orf81.pep	LIAVFFAFSIIANNVHYADYQSWMT									
	orf81a	LIAVFFAFSIIANNVHYAVYQSWITGINYWLMLKEITEVGGAGASMLDKLWLPALWGVLE									
		70	80	90	100	110	120				
		//									
	orf81.pep	QTVEFQLQKTPDGNWLPAYTSDHGQYVRQD									
	orf81a	IPHANGLEQISGGDIVDKYDNTIHKTDQMIOQTVEFQLQKQPDGNWLPAYTSDHGQYVRQD									
10		280	290	300	310	320	330				
	orf81.pep	IYNQGTVPQDSYLVPLVLYSPDKAVQQAANQAFAPCEIAFHQQLSTFLIHTLGYDMPVSG									
	orf81a	IYNQGTVPQDSYLVPLVLYSPDKAVQQAANQAFAPCEIAFHQQLSTFLIHTLGYDMPVSG									
15		340	350	360	370	380	390				
	orf81.pep	CREGSGVTGNLITGDAGSLNIRDGKAEYVYPQX									
	orf81a	CREGSGVTGNLITGDAGSLNIRDGKAEYVYPQX									
20		400	410	420							

The complete length ORF81a nucleotide sequence <SEQ ID 307> is:

25	1	ATGAAAAAT	CCCTTTTCGT	TCTCTTCTG	TATTCGGCC	TACTTACTGC
	51	CAGCGAAAT	GCTTATCCCT	TGTATTCCG	AATTGAACCC	TACCGGCTG
	101	CAAAAATGCG	AGAAACGTTT	GGCGTGACAT	TGTGATTATG	TGCGCTGTAT
	151	CTGTTTGCGC	GTTATAAGCG	AAACGCGTTG	TGTATTGCGG	TGTTTTTCGC
	201	GTTACGCAAT	ATTGCCAACA	ATGTGCAATTA	CGCGGTTTAT	CAAAGCTGGA
	251	TAACGGGCGAT	TAATTTATGG	CTGATGCTGA	AAGAGATTAC	CGAAGTTGGC
30	301	GGCGCAGGGG	CGTCGATGTT	GGATAAGTTG	TGGCTGCCTG	CGTTGTGGGG
	351	CGTGTGGGAA	GTCATGTTGT	TTTGACGCTT	TGCCAAGTTC	CGCCGTAAGA
	401	CGCATTTTTT	TGCCGATATA	CTGTTTGGCT	TCCTAATGCT	GATGATTTC
	451	GTGCGTTCGT	TGCACACGAA	ACAAGAACAC	GCTATTTCGC	CCAAACCGAC
35	501	ATACAGCCGC	ATCAAGGCCA	ATTATTTTCG	CTTCGGTTAT	TTTGTCCGAC
	551	CGCTGTTGCC	GTATCACTTG	TGTGATTAA	CGAGATGCC	TGCTGTCAA
	601	CAGCTGCTCT	CAGCACAAT	CGGCAAGCG	AGTATTCAAA	ATATCTCTCT
	651	GATTATGGGC	GAAAGCGAAA	GGCGCGGCGA	TTTGAATATG	TTTGCTTACG
	701	GGCGCGAAAC	TTGCGCGTTT	TTGACCCAGC	TTTGCAGACG	CGATTTTAAG
	751	CGGATTGTGA	AGCAAAGTTA	TTCCGCGAGC	TTTATGACGG	CAGTATCCCT
40	801	GCCAGTTTC	TTTAACTGTA	TACCGCATCG	CAACGGCTTG	GAACAAATCA
	851	GGCGCGGCGA	TATTGTGGAT	AAGTACGACA	ACACCATCCA	CAAACCGGAC
	901	CAATGATTCA	AAACCGTATT	CGAGCAGCTG	CAAAAGCAGC	CTGACGGCAA
	951	CTGGCTGTTT	GCCTATACCT	CGATCATATG	CGAATATGTT	CGCCAGAGTA
	1001	TCTACAATCA	AGGACGGGTG	CAGCCCGACA	GCTATCTCGT	CGCGCTGGTG
45	1051	TTGTACAGCC	CGGATAAGGC	CGTGCACAGC	GCTGCAGAAC	AGGCTTTTGC
	1101	GCCTTGCGAG	ATTGCGTTCC	ATCAGCAGCT	TTCAACGCTT	CTGATTACAA
	1151	CGTTGGGCTA	CGATATCCCG	GTTTCAGGTT	CTGCGCAAGG	CTCGCTAAGC
	1201	GGCAACCTGA	TTACGGGTGA	TGCAGCGAGC	TTGACATCTC	GCGACGGCAA
	1251	GGCGGAATAT	GTTTATCCCG	AATGA		

50 This encodes a protein having amino acid sequence <SEQ ID 308>:

55	1	MKKSFLVFLFL	YSSLLTASEI	AYRFVFGIET	LPAAKMAETF	ALTFTVIAALY
	51	LFARYKATRL	LIAVFFAFSI	IANNVHYAVY	QSWITGINYWL	MLKEITEVFG
	101	GAGASMLDKL	WLPALWGVLE	VMLFCSLAKF	RRKTHFSADI	LFALIMIMIF
	151	YRSFDTQKQH	GISPKPTYSR	IKANYFSFGY	FVGRVLPYQL	FDLSKIPVFK
	201	QPAPSRIGQC	SIQNVILVING	ESESAHLKL	FYGRGTSFPP	LTQLSQADFK
	251	PIVKQSYSGAC	FMIAVSLPSF	FNVIIPHANGL	EQISGGDIVD	KYDNTIHKTD
	301	QMIOQTVEFQL	QKQPDGNWLF	AYTSDHGQYV	RQDIYNGQTV	QPDYSLVPLV
	351	LYSPDKAVQQ	AANQAFAPCE	IAFHQQLSTF	LIHTLGYDMP	VSGCREGSGVT
	401	GNLITGDAGS	LNIRDGKAEY	VYPQ*		

60 ORF81a and ORF81-1 show 77.9% identity in 524 aa overlap:

		10	20	30	40	50	60
orf81a.pep	MKKSFLVFLYSSLLTASEIAYRFVFGIETLPAAKMAETFTLVIAALYLFARYKATRL						
	:						
orf81-1	MKKSFLVLVLYSSLLTASEIAYRFVFGIETLPAAKIAETFTLVIAALYLFARYKVTRL						
65		10	20	30	40	50	60

		70	80	90	100	110	120
	orf81a.pep	LIAVFFAFSIIANNVHYAVYQSWITG	INYLMLKEITEVGGAGASMLDKLWLPALWGVLE				
5	orf81-1	LIAVFFAFSIIANNVHYAVYQSWMTG	INYLMLKEVTEVGSAGASMLDKLWLPVWGVLE				
		70	80	90	100	110	120
	orf81a.pep	130	140	150	160	170	180
10	orf81a.pep	VMLFCSLAKFRKTHFSADILFAFLMLMIFVRSFDTQ	QHGHSIPKPTYSRIKANYFSFGY				
	orf81-1	VMLFCSLAKFRKTHFSADILFAFLMLMIFVRSFDTQ	QHGHSIPKPTYSRIKANYFSFGY				
		130	140	150	160	170	180
	orf81a.pep	190	200	210	220	230	240
15	orf81a.pep	FVGRVLPYQLFDLSKIPVFKQPAFSPRIQGGSIQ	INIVLIMGESESAHKLFGYGRSFPF				
	orf81-1	FVGRVLPYQLFDLSRIPAFKQPAFSPRIQGGSVQ	INIVLIMGESESAHKLFGYGRSFPF				
		190	200	210	220	230	240
	orf81a.pep	250	260	270	280		
20	orf81a.pep	LTQLSQADFKPIVKQSYSAGFMTAVSLSPSFNNVI	PHANGLEQISGGD-----				
	orf81-1	LTRLNQADFKPIVKQSYSAGFMTAVSLSPSFNAI	PHANGLEQISGGDTVMFLRAKEQGY				
		250	260	270	280	290	300
25	orf81a.pep	-----					
	orf81-1	TYFYSQAQENEMAILNLIGKKWIDHLIQFTQLGY	NGNDNMPDEKLLPLFDKINLQGGKHF				
30		310	320	330	340	350	360
	orf81a.pep	-----		290	300	310	320
	orf81-1	IVLHQSGSHAPYGALLQPDQKVFGEADIVDKY	DNTHIKTDQMIQTVFEQLQKQPDGNWLF				
35		370	380	390	400	410	420
	orf81a.pep	330	340	350	360	370	380
40	orf81a.pep	AYTSDHGQYVRQDIYNGQTVQPDPSYLVPLVLY	SPDKAVQQAANQAFAPCEIAFHQQLSTF				
	orf81-1	AYTSDHGQYVRQDIYNGQTVQPDPSYLVPLVLY	SPDKAVQQAANQAFAPCEIAFHQQLSTF				
		430	440	450	460	470	480
	orf81a.pep	390	400	410	420		
45	orf81a.pep	LIHTLGYDMPVSGCREGSVTGNLITGDAGSLN	IRDGKAEYVYPQX				
	orf81-1	LIHTLGYDMPVSGCREGSVTGNLITGDAGSLN	IRDGKAEYVYPQX				
		490	500	510	520		
50	<u>Homology with a predicted ORF from <i>N.gonorrhoeae</i></u>						
	The aligned aa sequences of ORF81 and a predicted ORF (ORF81.ng) from <i>N. gonorrhoeae</i> of the						
	N- and C-termini show 82.4 % and 97.5% identity in 85 and 121 overlap, respectively:						
	orf81.pep	MKKSFLTLVLYSSLLTASEIAYPLELGIETLPAAKIAETFALTIVIAALYLFARNKVTRL					60
55	orf81ng	MKKSFLVLFYSSLLTASEIAYRPFVGIETLPAARKAETFALTIFMIAALYLFARFKASRL					60
	orf81.pep	LIAVFFAFSIIANNVHYADYQSWMT					85
60	orf81ng	LIAVFFAFSMIANNVHYAVYQSWMTG	INYLMLKEVTEVGSAGASMLDKLWLPALWGVAE				120
	orf81.pep	//			QTVFEQLQKTPDGNWLFAYTSDHGQYVRQD		433
	orf81ng	ALLQPQDKVFGEADIVDKYDNTHIKTDQMIQTVFEQLQKQPDGNWLFAYTSDHGQYVRQD					433
65	orf81.pep	IYNGQTVQPDPSYLVPLVLYSPDKAVQQAANQAFAPCEIAFHQQLSTFLIHTLGYDMPVSG					493
	orf81ng	IYNGQTVQPDPSYIVPLVLYSPDKAVQQAANQAFAPCEIAFHQQLSTFLIHTLGYDMPVSG					493

		130	140	150	160	170	180
		190	200	210	220	230	240
5	orf81ng-1.pep	FVGRVLPYQLFDLSKI PVFKQAPAPSKIGGSIQNI VLMGESESAAHKLPGYGRGTSFP					
	orf81-1	FVGRVLPYQLFDLSRI PAFKQAPAPSKIGGSIQNI VLMGESESAAHKLPGYGRGTSFP					
		190	200	210	220	230	240
		250	260	270	280	290	300
10	orf81ng-1.pep	LTRLAQADFKPIVKQSYSGAGMTAVSLPFFNVI PHANGLEQISGGDTMMFRLAKEQGYE					
	orf81-1	LTRLAQADFKPIVKQSYSGAGMTAVSLPFFNVI PHANGLEQISGGDTMMFRLAKEQGYE					
		250	260	270	280	290	300
		310	320	330	340	350	360
15	orf81ng-1.pep	TYFYSAQAENQMAILNLNIGKKWIDHLIQPTQLGYGNGDNMPDEKLLPLFDKINLQGGHIF					
	orf81-1	TYFYSAQAENEMAILNLNIGKKWIDHLIQPTQLGYGNGDNMPDEKLLPLFDKINLQGGHIF					
		310	320	330	340	350	360
		370	380	390	400	410	420
20	orf81ng-1.pep	IVLHQRGSHAPYGALLQPDQKVFGEADIVDKYDNTIHKTDQMIQTVEFQLOKQPDGHWLF					
	orf81-1	IVLHQRGSHAPYGALLQPDQKVFGEADIVDKYDNTIHKTDQMIQTVEFQLOKQPDGHWLF					
		370	380	390	400	410	420
		430	440	450	460	470	480
25	orf81ng-1.pep	AYTSDHGGQYVRQDIYNQGTVPDYSYVPLVLYSPDKAVQQAANQAFAPCEIAFHQQLSTFF					
	orf81-1	AYTSDHGGQYVRQDIYNQGTVPDYSYVPLVLYSPDKAVQQAANQAFAPCEIAFHQQLSTFF					
		430	440	450	460	470	480
		490	500	510	520		
30	orf81ng-1.pep	LIHTLGYDMPVSGCREGSGVTGNLITGDAGSLNIRNGKAEYVYPQX					
	orf81-1	LIHTLGYDMPVSGCREGSGVTGNLITGDAGSLNIRNGKAEYVYPQX					
		490	500	510	520		

Furthermore, ORF81ng shows significant homology to an *E.coli* OMP:

40	gi 1256380 (U50906) outer membrane adherence protein-associated protein [E. coli] Length = 547 Score = 87.4 bits (213), Expect = 2e-16 Identities = 122/468 (26%), Positives = 198/468 (42%), Gaps = 70/468 (14%)	
45	Query: 25 VFGIETLPAAKMAETFA-LTFMIAALYLFARYKAS--RLLIIVFFAFPSMIANNVHYAVYQ 81 VFGI L A+ A L F + + + R + RLLA F + A + + ++Y Sbjct: 29 VFGITNLVASGGAHMVQRLLFFVLTLVVKRISSLPLRLLVAAFPVL-LTAADMSISLY- 86	
	Query: 82 SWMT-----GINYWLMLKEVTEVSGAGSMLDKLWLPALMGVAEVMFLCSLAKFRRT 134 SW T G ++ + EV A ML ++ P L A + L + + Sbjct: 87 SWCTFGTTFNDGFAISVLQSDPDEV---AKMLG-MYSPYLCAFAFLSLLFLVAIIRYDV 141	
50	Query: 135 HFSADILFAFLMLMIFVRSF-----DTKQEHGSIKPKPTYSRIKAN--YFSPGYVFG 183 + L+L++ S D K ++ SP SR +F+ YF Sbjct: 142 SLPTKVTGILLLLIVISGSLFSACQFAYKDAKNNAFSPYILASRFATYTPFFENLNYFAL 201	
55	Query: 184 RVLPYQ--LFDLSKI PVFKQAPAPSKIGGSIQNI VLMGESESAAHKLPGYGRGTSFPFL 241 +Q L + + P F + + I VLI+GES ++ L+GY R T+P + Sbjct: 202 AAKEHQRLLSIANTVYFQQL---SVRDTGIDTYVLIVGESVRVDNMSLYGYTRSTTPGV 257	
60	Query: 242 TRLSQADFKPIVKQSYSGAGMTAVSLP---SFFNVI PHANGLEQISGGDTMMFRLAKEQ 298 +G + Q+ S TFS+P + +V+ H I N+ A+ G Sbjct: 258 E--AQRKQIKLFNAISGAPYATLSVPLSLTADSVLSH-----DIHNYFDNIINMANQAG 310	
65	Query: 299 YETFYSAQA---ENQMAILNLNIGKKWIDHLIQPTQLGYGNGDNMPDEKLLPLFDKINLQ 355 ++T+ S+Q+ +N A+ ++ ++ + Y G DE LLP + Q Sbjct: 311 FQTFWLSSQAFRNGTAVTSI-----AKRAMETVYVRGF---DELLPHLSQALQ 359	
70	Query: 356 --QGRHFIVLHQRGSHAPYGALLQPDQKVFGEADIVDK-YDNTIHKTDQMIQTVEFQLOK 412 Q + IVLH GSH P + VF D D YDN+H TD ++ VFE L+ Sbjct: 360 NTQQKKLIVLHNGSIEFACSAFYQSSAVFPQDDQDQACYNISHTYDLSLQGVFELLK- 418	

Query: 413 QPQDNWLFAYTSDHG---QYVRQDIYNQO--TVQPDYSYIVPL-VLYSP 454
 D Y +DHG +++Y G +Y VP+ + YSP
 Sbjct: 419 --DRRASVMYFADHGLERDPTKKNVYFHGGREASQQAHYHFMFWYSP 464

5

Based on this analysis, including the presence of a putative leader sequence (double-underlined) and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

10 Example 37

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 311>:

```

1    ...ACCTGCTCC TCTTCATGCC CCTCGTCTC ACAC GTGCG GCACACTGAC
51   CGGCATACTC GCCACGGCG GCGGCAACCG CTTTGCGCTC GAACAAGAAC
101  TCGTCCGCGC ATCGTCCGCG GCGCGCGTCA AAGAAATGGA TTTGTCCGCG
151  YTAAGAAGAC GCAAGACGCG CyTTTACGTC TCGGTATATG GCGACCAAGG
201  TTCCGGCAAC ATAAGCGGCG GACGCTACTC TATCGAGCCA CTGATACCGG
251  GCGGCTACCA CAACAACGCC GAAGTGCACA CCAATACAG CTACCCCGCC
301  TACGCACTAC CCGCCACCA CAAATCCGAC GCGCTCTCCA GCGTAACCA
351  TTCCACATCG CTTTTGAACG CCGCGGCGCG CGYCYTAGCA AAAAAGACGC
20  401  GACGCAAGG CGAAGCGTCC CCGCGACTGT CCGTCAACGG CAGCGGCGAC
451  TACCGCAACG AAACCGTCTG CGCAACCGCC CGGACGCTTT CTTCTCTGAC
501  CAACCTCATC CAAACCGTCT TCTACCTCGG CGGCATCGAA GTGgTACCGC
551  CCGTATACCG CGACACGCGC GTATTCTGTA CCGTGCAGCT A...
```

This corresponds to the amino acid sequence <SEQ ID 312; ORF83>:

```

25  1    ..TLLLEPIPLV TXCGTLTGIL AHGGGKRFAY EQELVAASSR AAVKEMDLA
51   LKGRKAAXYV SVMGDQSGN ISGGYRYSIDA LIRGGYHNNP ESATQYSYPA
101  YDTATTTKSD ALSVTTSTS LLNAPAXLT KNSGRKGRS AGLSVNGTGD
151  YRNETLLANP RDVSLTNLI QTVFYLRLGIE VVPPXYADTD VFVTVDV...
```

Further work revealed the complete nucleotide sequence <SEQ ID 313>:

```

30  1    ATGAAAACCC TGCTCCTCCT CATCCGCCCT GTCTCAGAC CCGCGGCGAC
51   ACTGACCGCG ATACCGCGCC ACGGCGGCG CAAACGCTTT GCCGTGCAAC
101  AAGAAGCTGT CGCGCATCG TCCGCGCGCG CCGTCAAGA AATGATTGTT
151  TCGCGCCTAA AAGGACGCAA AGCGCGCCTT TAAGTCTCCG TTATGGGCGA
201  CCAAGGTTTC GGCAACATAA GCGGCGGAGC CTACTCTATC GACGCACTGA
35  251  TACGCGCGCG CTACCAACAC AACCOCGAAA GTGCGCACCA ATACAGCTACT
301  CCGCGCTACG ACATACCGC CACCAACAAA TCCGACGCG CTCTCAGCT
351  AACCACTTCC ACATCGCTTT TGAACGCCCG CCGCGCGCGC CTGACCAAAA
401  ACAGCGGACG CAAAGCGGAA CGTCCGCCCG GACTGTCTGT CAACGCGACG
451  GCGCACTACC GCAACGAAC CCGTCTCTCA AACCOCGCG ACOTTTCTTT
40  501  CCGTACCAAC CTCATCAAAA CCGTCTCTTA CCGCGCGCGC ATCGAAGTCG
551  TACCGCGCGA ATACCGCGAC ACCGACGTAT TCGTAACGCT CGAGCTATTC
601  GCGACCGTCC GCAACCGTAC CGAAGTGCAC CTCTACAACG CCGAAACCTT
651  TAAAGCCCAA ACCAAGCTCG AATATTTCGC CGTTGACCGC GACAGCGGAA
45  701  AACTGCTGAT TACCCCTAAA ACCGCGCGCT ACGAATCCCA ATACCAAGAA
751  CAATACGCCC TTGGAACCGG CCTTACAAA GTACAGCAAAA CCGTCAAGAC
801  CTCAGACCGC CTGATGGTTC ATTCTCCGCA CATTACCCCG TACGGGCGAC
851  CAACCGCGCA AAACCGTCCC GACTTCAAAC AAAACAACGG TAAAAAACCC
901  GATGTGCGCA ACGAAGTCAT CCGCGCGCGC AAAGGAGGAT AA
```

This corresponds to the amino acid sequence <SEQ ID 314; ORF83-1>:

```

50  1    MKTLLLLLPL VLTACTLGTG IPAHGGGKRF AVEQELVAAS SRAAVKEMDL
51   SALKGRKAAL YSVVMGDQGS GNISGGYRYS DALIRGGYHN NPSATQYSY
101  PAYDTTATTK SDALSVTTTS TSLNAPAAA LTKNSGRKGE RSAGLSVNGT
151  GDYRNETLLA NPDVSLFTN LIQTVFYLRL IEVVPPEYAD TDFVTVVDVF
201  GTVRSRTELH LYNAETLKAQ TKLEYFAVDR DSRKLLITPK TAAYESQYOE
55  251  QYALWTGPYK VSKTVKASDR LMVDFSDITP YGDTTAQNRP DFKQNKNGKK
```


301 DVGNEVIRRR KGG*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)ORF83 shows 96.4% identity over a 197aa overlap with an ORF (ORF83a) from strain A of *N.*5 *meningitidis*:

		10	20	30	40	50
orf83.pep		TLLLFPLVLTCGTLTGIPAHGGGKRF	AVEQELVAASSRAAVKEMDLS	SALKGRKAAX		
10	orf83a	MKTLKLLIPLVLTACGTLTGIPAHGGGKRF	AVEQELVAASSRAAVKEMDLS	SALKGRKAAL		
		60	70	80	90	100
orf83.pep		YVSMGDSQSGNISGGGRYSIDALIRGGYHNNPESATQYSYPAYDTTATKSDALSSVTT				
15	orf83a	YVSMGDSQSGNISGGGRYSIDALIRGGYHNNPESATQYSYPAYDTTATKSDALSSVTT				
		120	130	140	150	160
orf83.pep		TSLLNAPAAKLTKNSSGRKGRSAGLSVNGTGDYRNETLLANPRDVSFLTNLIQTVEYL				
20	orf83a	TSLLNAPAAKLTKNSSGRKGRSAGLSVNGTGDYRNETLLANPRDVSFLTNLIQTVEYL				
		180	190			
orf83.pep		IEVVPPEYADTVDFVTVDFVGT				
25	orf83a	IEVVPPEYADTVDFVTVDFVGT				

The complete length ORF83a nucleotide sequence <SEQ ID 315> is:

30	1	ATGAAACCC	TGCTGTCCT	CATCCGCTC	GTCCGACAG	CCTCGGGCAC
	51	ACTGACCGGC	ATACCGCGCC	ACGGCGCGG	CAAAGCGTT	GCGTCCGAC
	101	AAGAAGCTGT	CGCGCATCG	TCCCGCGCG	CGCTCAAAG	AATGGACTTG
	151	TCCGCGCTGA	AAGGACGCA	AGCGCGCCT	TACGCTCTCG	TTATGGGCGA
	201	CCAAGGTTTG	GGCAACATA	GCGCGGAGC	C7ACTCTATC	GACGCACTGA
35	251	TACGCGGCG	CTACACAAAC	AACCGCGAA	TCCGACCGCA	ATACAGCTAC
	301	CCGCGCTACG	ACACTACGC	CACCAACAA	TCCGACCGGC	TCTCCAGCGT
	351	AACCACTTC	ACATCGCTT	TGAACGCGC	CGCGCGCGC	CTGACGAAAA
	401	ACAGCGGAG	CAAAGCGAA	CGCTCGCGC	GACTCTCGCT	CAACGGCACG
	451	GCGGACTACC	GCAAGCAAA	CCTGCTCGC	AACCGCGCG	AGCTTCTCTT
40	501	CCTGACCAAC	CTCATCCAA	CGCTCTCTA	CCTGGGCGG	ATCGAAGTCG
	551	TACGCGCGA	ATACGCGAC	ACGAGCGAT	TCTGTACCGT	CGAGTATTTC
	601	GGACCGCTC	GCAGCGCAC	CGAAGCTCG	CTGTACACG	CGCAACCTC
	651	TAAAGCCCA	ACCAAGCTG	ATAATTTCG	CGTGAACCG	GACAGCGCGA
	701	AAGTCTGAT	TGCCCTTAA	ACGCGCGCT	ACGAATCCA	ATACCAAGAA
45	751	CAATACGCG	TCTGGATGG	ACCTTACAG	GTCCGCAAAA	CGCTCAAGC
	801	CTCAGACCG	CTGATGTCG	ATTCTCTCG	CATCAACCC	TACGCGGCA
	851	CAACCGCCA	AAACCGTCC	GACTTCAAC	AAAACAACG	TAAAAAACC
	901	GATGTCGCA	ACGAAGTCAT	CGCGCGCGC	AAAGAGGAT	AA

This encodes a protein having amino acid sequence <SEQ ID 316>:

50	1	MKTLKLLIPL	VLTACGTLTG	IPAHGGGKRF	AVEQELVAAS	SRAAVKEMDL
	51	SALKGRKAAL	YVSMGDSQGS	GNISGGRYST	DALIRGGYHN	NPESATQYSY
	101	PAYDTTATTK	SDALSSVTT	TSLLNAPAAA	LTKNSGRKGE	RSAGLSVNGT
	151	GDYRNETLLA	NPRDVSFLTN	LQITVEYL	IEVVPPEYAD	TDFVTVDFV
	201	GTVRSRTELH	LYNAETLKAQ	TKEYFAVDR	DSRKLLIAFK	TAYEYQYQE
55	251	QYALWMGPGS	VGKTVKASDR	LMVDVSDTF	YGDTTANRNP	DFKNGKGGK
	301	DVGNEVIRRR	KGG*			

ORF83a and ORF83-1 show 98.4% identity in 313 aa overlap:

		10	20	30	40	50	60
orf83a.pep		MKTLKLLIPLVLTACGTLTGIPAHGGGKRF	AVEQELVAASSRAAVKEMDLS	SALKGRKAAL			

ORF83 shows 94.9% identity over a 197aa overlap with a predicted ORF (ORF83.ng) from *N. gonorrhoeae*:

	orf83.pep	TLILLFIPVLVTGCTGTGLANGGGKRFVFEQLVAASSRAAVKEMDLSALGKRKAAX	58
40	orf83.ng	MTLTLILLFIPVLVTAGCTGTGPAHGGGGKRFVFEQLVAASSRAAVKEMDLSALGKRKAAL	60
	orf83.pep	YVSMVGDDGSGNISGGRYSIDALIRGGYHNHPESATQYSYPAYDTTATTKSDALS SVTTS	118
45	orf83.ng	YVSMVGDDGSGNISGGRYSIDALIRGGYHNHPESATRYSPAYDTTATTKSDALSGVTTS	120
	orf83.pep	TSLLNPAAXLTKNSGRKGERSAGLSVMGTGDYRNETLLANPRDVSFLTNLIQTVFYLRG	178
	orf83.ng	TSLLNPAAXLTKNKGKRGERSAGLSVMGTGDYRNETLLANPRDVSFLTNLIQTVFYLRG	180
50	orf83.pep	IEVVPKYADTVFTVVDV	197
	orf83.ng	IEVVPPEYADTVFTVVDVFGTGRSRTLEHLYNAETLKAOTKLEYFAVDRDSRKLIIAPK	240

The complete length ORF83ng nucleotide sequence <SEO ID 317> is:

55	1	A	GAAAMCCC	TGCTCTCTCT	CATCTCCCCCT	GTACTCACCG	CCTGGGGCAC
	51	A	CTGACCGCG	ATACCGGCC	ACGGCGGGCT	CAAGCGCTTT	CGCCTCGAAC
	101	A	GGAATCTCT	CGCGCATCG	TCCCGCGCGC	CCGCTAAGA	AATGGAATTG
	151	T	CGCGCCTGA	AAGAACGAA	AGCGCGCTTT	TACCTTCCG	TTATGGGCGA
	201	C	CAAGTTGTC	GGCAACATA	CGCGGCGGCG	CTACTCTATC	GACGCACTGA
	251	T	ACGCGGGCG	CTACACAAC	AACCCGACGA	CGCGCCACCG	ATACAGCTAG
60	301	C	CGCGCTATG	ACACTACCG	CACCAACAAA	TCCGACGGCG	TCCTCGGCGT
	351	A	ACCACTTCT	ACATCGTTT	TGARCGCGCC	CGCGCGCGCG	CTGACAGAAA
	401	A	CAACGGGAC	CAAGGGCGA	AGCTCTCGCG	GACTGTCGCT	CACGCGGACG
	451	G	CGCGATGCT	CGCGCGGCG	CTCTGCTGGC	ACGCGCTCTT	AGCGGCTCTT
	501	C	CTGACGAC	CTCATCCAA	CGCTCTCTA	CTCGTACGCT	ATCGAAGGCT
65	551	T	ACCGGCCGA	ATACGGCGA	ACCGACGATAT	TGCTGATACG	CGACGTATTC
	601	G	GCAACGCTC	GCAGCGGTAC	CTGATCGAC	CTCTACACG	CGAAACCCCT

5
 651 TAAAGGCCAA ACCAAGCTCG AATATTTGCG CGTCGACGCG GACAGCCGGA
 701 AACTGCTGAT TGCCCTCTAAA ACCGCGCGCT ACGAATCCCA ATACCAAGAA
 751 CAATACGCCC TCTGGATGGG ACCTTACAGC GTCGGCAAAA CCGTCAAGGC
 801 CTCAGACGCG CTGATGGTGG ATTTCTCCGA CATCACCCCC TACGGCGACA
 851 CAACCGGCCA AAACCGTCCC GACTTCARAC AAAACAACCG TAAAAACCCC
 901 GATGTCGGCA ACGAAGTCAT CGCCGCGCG AAAGGAGGAT AA

This encodes a protein having amino acid sequence <SEQ ID 318>:

10
 1 M K T L L L L I P L V L T A C G T L T G I P A H G G G K R F A V E Q E L V A A S S R A A V K E M D L S A L K G R K A A L
 51 S A L A G R K A A L Y V S V M G D Q G S G N I S G G R Y S I D A L I R G G T H N N P D S A T R Y S Y
 101 F A Y D T T A T T K S D A L S G V T T S T S L N A P A A A L T K N G R K G E R S A G L S V N G T
 151 G D Y R N E T L L A N P R D V S F L T N L I Q T V F Y L R G I E V V P P E Y A D T D V F V T V D F
 201 G T V R S R T E L H L Y N A E T L K A Q T K L E Y F A V D R D S R K L L I A P K T A A Y E S Q Y Q E
 251 Q Y A L W M G P Y S V G K T V K A S D R L M V D F S D I T P Y G D T T A Q N R P D F K Q N N G K N P
 301 D V G N E V I R R K G G *

15 ORF83ng and ORF83-1 show 97.1% identity in 313 aa overlap

		10	20	30	40	50	60
orf83-1.pep		M K T L L L I P L V L T A C G T L T G I P A H G G G K R F A V E Q E L V A A S S R A A V K E M D L S A L K G R K A A L					
orf83ng		M K T L L L I P L V L T A C G T L T G I P A H G G G K R F A V E Q E L V A A S S R A A V K E M D L S A L K G R K A A L					
		10	20	30	40	50	60
		70	80	90	100	110	120
orf83-1.pep		Y V S V M G D Q G S G N I S G G R Y S I D A L I R G G Y H N N P E S A T Q Y S Y P A Y D T T A T T K S D A L S V T T S					
orf83ng		Y V S V M G D Q G S G N I S G G R Y S I D A L I R G G Y H N N P E S A T R Y S Y P A Y D T T A T T K S D A L S V T T S					
		70	80	90	100	110	120
		130	140	150	160	170	180
orf83-1.pep		T S L N A P A A A L T K N S G R K G E R S A G L S V N G T G D Y R N E T L L A N F R D V S F L T N L I Q T V F Y L R G					
orf83ng		T S L N A P A A A L T K N N G R K G E R S A G L S V N G T G D Y R N E T L L A N F R D V S F L T N L I Q T V F Y L R G					
		130	140	150	160	170	180
		190	200	210	220	230	240
orf83-1.pep		I E V V P P E Y A D T V F V T V D V F G T V R S R T E L H L Y N A E T L K A Q T K L E Y F A V D R D S R K L L I T P K					
orf83ng		I E V V P P E Y A D T V F V T V D V F G T V R S R T E L H L Y N A E T L K A Q T K L E Y F A V D R D S R K L L I A P K					
		190	200	210	220	230	240
		250	260	270	280	290	300
orf83-1.pep		T A A Y E S Q Y Q E Q Y A L W M G P Y S V G K T V K A S D R L M V D F S D I T P Y G D T T A Q N R P D F K Q N N G K N P					
orf83ng		T A A Y E S Q Y Q E Q Y A L W M G P Y S V G K T V K A S D R L M V D F S D I T P Y G D T T A Q N R P D F K Q N N G K N P					
		250	260	270	280	290	300
		310					
orf83-1.pep		D V G N E V I R R K G G X					
orf83ng		D V G N E V I R R K G G X					
		310					

Based on this analysis, including the presence of a putative ATP/GTP-binding site motif A (P-loop) in the gonococcal protein (double-underlined) and a putative prokaryotic membrane lipoprotein lipid attachment site (single-underlined), it is predicted that the proteins from *N. meningitidis* and *N. gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 38

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID 319>:

```

1  ATGGCAGAGA TCTGTTTGAT AACCGGCACG CCGGGTTACG GGAAACATT
5  51  AAAAATGGTT TCCATGATGG CGAATGATGA AATGTTTAAG CCTGATGAAA
101 AAGCCATACG CGGTAAAGTA TTTACGAACA TAAAGGGCTT GAAATACCG
151 CACACCTACA TAGAAACCGA CGCAAAAAGC CTGCCGAAT CGACAGATGA
201 GCAGCTTTCG GCGCATGATA TGTACGAATG GATAAAGAAC CCCGAAATA
251 TCGGGTCTAT TGTCAATTGA GATGAAGCTC AAGACGTATG GCCGCGACGC
10  301 TCGCAGGGTT CAAAATCCC TGAAAATGTC CAATGGTGTA ATACGCACAG
351 ACATCAGGGC ATTGATATAT TTGTTTTCAC TCAAGTCCCT AAGCTTCTAG
401 ATCAAAATCT TAGAACGCTT GTACGGAAC ATTACCACT CGCTTCAAAC
451 AAGATGGGTA TGCCTAGCCT TTTAGATGG AAAATATGCG CGACGATCC
501 CGTAAAAATG GCATCAAGCG CATCTCCAG TATCTATACA CTGGATAAAA
15  551 AAGTTTATGA CTTGTAGTAT TmCGCGAAG TTCAATCCGT AAATAGGTC
601 AAGCGGTCAA AGTGGTTTTA CACTCTGCCA GTAATAGTAT TGCTGATTCC
651 CGTGTTCGTC GGCCTGTCCT ATAAAAATGT GgCaGTTAC GGAAAAAACC
701 aGGAAGAACC CGCAGCACAA GAATCGCGCG CAACAGAACA GCAGGCAGTA
751 CTTCCGGATA AAACAGAAGC CGAGCCGGTA AATAACGGCA ACCTTACCGC
20  801 AGATATGTTT GTTCCGACAT TGTCCGAaAA ACCCGAAGC AAGCGaTTT
851 ATAACGGTGT AAGCGAGTA AGAACCTTTG AATATATAGC AGGCTGTATA
901 GAAGCGGAA GAACCGGATG CGCCTGCTAT TCGCaTCAAG GGACGGCATT
951 gaAAGAAGTG ACGGaGTTGA TGTGCGaAgG aCTATGTAA AaACGGCTTG
25  1001 CGCTTTAACC CATCAAAAGA AGAAAGCCAA GCGCGGAAAG TTCAACGAAG
1051 CGCGCAGAAC CATTCGTCCT GCGCGCAAG TTCCCAATT GCGCGGAAA
1101 CGGTAGCAGA ACCTAATGTA CGATAATTGG GAAGAAGCGG GAAACCGTT
1151 TGAAGGAATC GGaCGGGGCG GTCGTCGATC CGGCAAACTG A

```

This corresponds to the amino acid sequence <SEQ ID 320; ORF84>:

```

1  MAEICLITGT PGSGKTLKMW SMMADEMFK PDEKAIRRV FTNIKGLKIP
30  51  HTYIETDAKK LFKSTDEQLS AHDMYEWIK PENIGSIIV DEAQDVWPAR
101 SAGSKIPENV QWLNTHRRQG IDIFVLTPGP KLLDQNLRL VRKHVHIAN
151 KGMNRTLLEW KICADDPVKM ASSAFSSIYT LOKKVYDLYX XAEVHTVNRV
201 KRSKWFYTLF VIVLLIPVFV GLSYKMLSSY GKQVEEPAQQ ESARTEQQAV
251 LPDKTEGEPV NNGNLIDAMF VPTLSEKPSX KPIYNGVRQV RTEYIAGCI
35  301 EGGRTGCACY SHQQTALKEV TELMCKDYVK NGLPFNPKYE ESQGGVEQQS
351 AQHSDRAQV ATLGGKPKQN LMYDNWEERG KPFEIGGGGV VGSAN*

```

Further work revealed the complete nucleotide sequence <SEQ ID 321>:

```

1  ATGGCAGAGA TCTGTTTGAT AACCGGCACG CCGGGTTACG GGAAACATT
40  51  AAAAATGGTT TCCATGATGG CGAATGATGA AATGTTTAAG CCTGATGAAA
101 ACGGCATACG CGGTAAAGTA TTTACGAACA TAAAGGGCTT GAAATACCG
151 CACACCTACA TAGAAACCGA CGCAAAAAGC CTGCCGAAT CGACAGATGA
201 GCAGCTTTCG GCGCATGATA TGTACGAATG GATAAAGAAC CCCGAAATA
251 TCGGGTCTAT TGTCAATTGA GATGAAGCTC AAGACGTATG GCCGCGACGC
45  301 TCGCAGGGTT CAAAATCCC TGAAAATGTC CAATGGTGTA ATACGCACAG
351 ACATCAGGGC ATTGATATAT TTGTTTTCAC TCAAGTCCCT AAGCTTCTAG
401 ATCAAAATCT TAGAACGCTT GTACGGAAC ATTACCACT CGCTTCAAAC
451 AAGATGGGTA TGCCTAGCCT TTTAGATGG AAAATATGCG CGACGATCC
501 CGTAAAAATG GCATCAAGCG CATCTCCAG TATCTATACA CTGGATAAAA
55  551 AAGTTTATGA CTTGTACGAA TCAGCGGAAG TTCAATCCGT AAATAGGTC
601 AAGCGGTCAA AGTGGTTTTA CACTCTGCCA GTAATAGTAT TGCTGATTCC
651 CGTGTTCGTC GGCCTGTCCT ATAAAAATGT GAGCAGTTAC GGAAAAAACC
701 AAGGAAGAACC CGCAGCACAA GAATCGCGCG CAACAGAACA GCAGGCAGTA
751 CTTCCGGATA AAACAGAAGC CGAGCCGGTA AATAACGGCA ACCTTACCGC
801 AGATATGTTT GTTCCGACAT TGTCCGAaAA ACCCGAAGC AAGCGATTT
851 ATAACGGTGT AAGCGAGTA AGAACCTTTG AATATATAGC AGGCTGTATA
901 GAAGCGGAA GAACCGGATG CGCCTGCTAT TCGCATCAAG GGACGGCATT
951 GAAGAAGTG ACGGaGTTGA TGTGCAAGGA CTATGTAAAA AACGGCTTCG
1001 CGTTTAAACC ATACAAAGAA GAAGACCAAG GGCAGGAAGT TCAGCAAGC
1051 GCGCAGCAAC ATTCGCGACG GCGCGCAAGT GCCACATTGG GCGGAAACC
60  1101 GTACGAGAAC CTAATGTACG AATAATTGGG AGAACGCGGG AAACCGTTG
1151 AAGGAATCGG CGGGGGCGTG GTCGATTCCG CAAACTGA

```

This corresponds to the amino acid sequence <SEQ ID 322; ORF84-1>:

1 MAEICLITGT PGSGKTLKQV SMMADEMFK PDENGIRKRV FTNKGKLP
 51 HTYIETDAKK LPKSTDEQLS AHMVEWIKK PENIGSIVIV DEAQDWWFAR
 101 SAGSKIPENV QWLNTHRRQG IDIFVLTOGP KLDDQNLRTL VRKHVHTASN
 151 KGMRTLLLEW KICADDPVKM ASSAFSSIY LOKKVYDLYE SAEVHTVNVK
 201 KRKWFYTLF VIVLLIPVFL GLSYKMLSSY GKQKEEPAQ ESAATEQQA
 251 LPDKTEGEPV NNGNLTADMV VPTLSEKPS KPIYNGVRQV RTFEYIAGCI
 301 EGGRTGCACY SHQGTALKEV TELMCKDYVK NGLFPNPKYK ESQGEVQQS
 351 AQHSDRAQV ATLGKPK*QN LMYDNWEERG KPFEIGGGV VGSAN*

Computer analysis of this amino acid sequence gave the following results:

10 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF84 shows 93.9% identity over a 395aa overlap with an ORF (ORF84a) from strain A of *N.*

meningitidis:

		10	20	30	40	50	60
15	orf84.pep	MAEICLITGT	PGSGKTLKQV	SMMADEMFK	PDENGIRKRV	FTNKGKLP	HTYIETDAKK
	orf84a	MAEICLITGT	PGSGKTLKQV	SMMADEMFK	PDENGIRKRV	FTNKGKLP	HTYIETDAKK
		10	20	30	40	50	60
20	orf84.pep	LPKSTDEQLS	AHMYEWIKK	PENIGSIVIV	DEAQDWWFAR	SAGSKIPENV	QWLNTHRRQG
	orf84a	LPKSTDEQLS	AHMYEWIKK	PENIGSIVIV	DEAQDWWFAR	SAGSKIPENV	QWLNTHRRQG
		70	80	90	100	110	120
25	orf84.pep	IDIFVLTOGP	KLDDQNLRTL	VRKHVHTASN	KGMRTLLLEW	KICADDPVKM	ASSAFSSIY
	orf84a	IDIFVLTOGP	KLDDQNLRTL	VRKHVHTASN	KGMRTLLLEW	KICADDPVKM	ASSAFSSIY
		130	140	150	160	170	180
30	orf84.pep	LOKKVYDLYE	SAEVHTVNVK	KRSKWFFYLE	VIVLLIPVFL	GLSYKMLSSY	GKQKEEPAQ
	orf84a	LOKKVYDLYE	SAEVHTVNVK	KRSKWFFYLE	VIVLLIPVFL	GLSYKMLSSY	GKQKEEPAQ
		190	200	210	220	230	240
35	orf84.pep	ESAATEQQA	LVPKTEGEPV	NNGNLTADMV	VPTLSEKPS	KPIYNGVRQV	RTFEYIAGCI
	orf84a	ESAATEQQA	LVPKTEGEPV	NNGNLTADMV	VPTLSEKPS	KPIYNGVRQV	RTFEYIAGCI
		250	260	270	280	290	300
40	orf84.pep	EGGRTGCACY	SHQGTALKEV	TELMCKDYVK	NGLFPNPKYK	ESQGEVQQS	AQHSDRAQV
	orf84a	EGGRTGCACY	SHQGTALKEV	TELMCKDYVK	NGLFPNPKYK	ESQGEVQQS	AQHSDRAQV
		310	320	330	340	350	360
45	orf84.pep	ATLGKPKQNL	MYDNWEERG	KPFEIGGGV	VGSANX		
	orf84a	ATLGKPKQNL	MYDNWEERG	KPFEIGGGV	VGSANX		
		370	380	390			

The complete length ORF84a nucleotide sequence <SEQ ID 323> is:

55	1	ATGCGACAGA	TCTGTTTGAT	AACCGGCACG	CCCGGTTTAC	GGAAACATT
	51	AAAAATGGTT	TCCATGATG	CAACGATAGA	ATGTTTAAAG	CCGATGAA
	101	ACGCGATACG	CCGTAAACFA	TTTACGAAAG	TCAATGGCTT	GAGATACCG
	151	CACACCTACA	TAGAAACGGA	CGCGAAAG	CTGCCGAAT	CGACAGATGA
	201	GCAGCTTTCG	GCGCATGATA	TGTACGAATG	GATAAAGAG	CCCGAAATTA
60	251	TCGGGTCTAT	TGTCATATGA	GATGAAGCTC	AAGCATATG	GC CGCGACGC
	301	TCGGCAGGTT	CAAAATATCC	TGAAATATGC	CAATGGCTGA	ATACGCACAG
	351	ACATCAGGGC	ATTGATATAT	TTGTTTIGAC	TCAAGGCTCT	AAGCTTCTAG
	401	ATCAAAATCT	TAGAACGCTT	GTACGGAAC	ATTACCATAT	CGCTTCAAC
	451	AAGATGGGTA	TGCGTACGCT	TTAGAATGG	AAATATGCG	CGGACGATCC

501 CGTAAAAATG GCATCAAGCG CATTCTCCAG TATCTATACA CTGGATAAAA
 551 AAGTTTATGA CTTGTACGAA TCAGCGGAAG TTCATACCGT AAATAAGGTC
 601 AAGCGGTCAA AATGGTTTTA TACTCTGCCA GTAATAATAT TGCTGATPTC
 651 CGTTTTTGTC GGCTCTGCTT ATAAATATGT AGTAGTATAT GGAATAAATC
 701 AGGAAGAACC CSCAGACAAA GAATCGGCGG CACACAGACA TCAGGCAGTA
 751 TTTTCAGGATA AATCAGGAGG CGAGCCGGTA AACACAGGTA ACCTTACCAG
 801 AGATATGTTT GTTCCGACAT TGTCCGAAAA ACCGGAAGG AAGCCGATT
 851 ATAACGGTGT AAGCGAGGTA AGAACCTTTG AATATATAGC AGGCTGTGTA
 901 GAAGGCGGAA GAACCGGATG CACATGCTAT TCGCATCAAG GGACGGCAT
 951 GAAAGAAATT ACAAGAGAAA TGTGCAAGGA TTACGCAAGA AACGGATTGC
 1001 CGTTTAAACC ATATAAAGAA GAAAGCCAAG GCGCGGATGT CAGCAAAAGT
 1051 GAGCAGCACC ATTCCGACAG ACCGCAAGTT GCGCAGTTGG CGCGAAAGCC
 1101 GTGGCAAAAT CTATGTATG ATAATTGACA GGAGCGGAGA AAACCGTTTG
 1151 AAGGAATCGG CGGGGGCGTG GTCCGATCGG CAAACTGA

15 This encodes a protein having amino acid sequence <SEQ ID 324>:

1 MAEICLITGT PGSGKTLKMW SMANDEMFK PDENGIRRVK FTKNIKGLKIP
 51 HTYIETDAKK LPKSTDEQLS AHDMYEWIKK PENIGSIVIV DEAQDVWPAR
 101 SAGSKI PENV QWLNTHRHOG IDIFVL TQGS KLLDQNLRL VRKHVHIAS
 151 KGMRLTLEW KICADDFVKK ASSAFSSIYT LDKKYVDLYE SAEVHTVNVK
 201 KRKSKWFTLP VILLIPVFEV GLSKMLSSY GKQEEPAQ ESAATEHQAV
 251 FDKRTGEQPV NNGNLADMF VPTLSEKPEK KPIYNGVRQV RTFEYIAGCV
 301 EGGRTGCTCY SHQGTALKEI TKEMCKDYAR NGLPFPNPKY ESQGRDVQQS
 351 EQHHSRDPQV ATLGKPKWQN LMYDNWQERG KPEGIGGGV VGSAN*

ORF84a and ORF84-1 show 95.2% identity in 395 aa overlap:

25		10	20	30	40	50	60
	orf84a.pep	MAEICLITGT	PGSGKTLKMW	SMANDEMFK	PDENGIRRVK	FTKNIKGLKIP	HTYIETDAKK
	orf84-1	MAEICLITGT	PGSGKTLKMW	SMANDEMFK	PDENGIRRVK	FTKNIKGLKIP	HTYIETDAKK
30		70	80	90	100	110	120
	orf84a.pep	LPKSTDEQLS	AHDMYEWIKK	PENIGSIVIV	DEAQDVWPAR	SAGSKI PENV	QWLNTHRHOG
	orf84-1	LPKSTDEQLS	AHDMYEWIKK	PENIGSIVIV	DEAQDVWPAR	SAGSKI PENV	QWLNTHRHOG
35		70	80	90	100	110	120
	orf84a.pep	IDIFVL TQGS	KLLDQNLRL	VRKHVHIAS	NKGMRLTLEW	KICADDFVKK	ASSAFSSIYT
	orf84-1	IDIFVL TQGS	KLLDQNLRL	VRKHVHIAS	NKGMRLTLEW	KICADDFVKK	ASSAFSSIYT
40		130	140	150	160	170	180
	orf84a.pep	LDKKYVDLYE	SAEVHTVNVK	VKRKSWFTLP	VILLIPVFEV	GLSKMLSSY	GKQEEPAQ
	orf84-1	LDKKYVDLYE	SAEVHTVNVK	VKRKSWFTLP	VILLIPVFEV	GLSKMLSSY	GKQEEPAQ
45		190	200	210	220	230	240
	orf84a.pep	ESAEATEHQAV	FDKRTGEQPV	NNGNLADMF	VPTLSEKPEK	KPIYNGVRQV	RTFEYIAGCV
	orf84-1	ESAEATEHQAV	FDKRTGEQPV	NNGNLADMF	VPTLSEKPEK	KPIYNGVRQV	RTFEYIAGCV
50		250	260	270	280	290	300
	orf84a.pep	EGGRTGCTCY	SHQGTALKEI	TKEMCKDYAR	NGLPFPNPKY	ESQGRDVQQS	EQHHSRDPQV
	orf84-1	EGGRTGCTCY	SHQGTALKEI	TKEMCKDYAR	NGLPFPNPKY	ESQGRDVQQS	EQHHSRDPQV
55		310	320	330	340	350	360
	orf84a.pep	ATLGKPKWQN	LMYDNWQERG	KPEGIGGGV	VGSANX		
	orf84-1	ATLGKPKWQN	LMYDNWQERG	KPEGIGGGV	VGSANX		
60		370	380	390			
	orf84a.pep	ATLGKPKWQN	LMYDNWQERG	KPEGIGGGV	VGSANX		
	orf84-1	ATLGKPKWQN	LMYDNWQERG	KPEGIGGGV	VGSANX		
65		370	380	390			

Homology with a predicted ORF from *N.gonorrhoeae*

ORF84 shows 94.2% identity over a 395aa overlap with a predicted ORF (ORF84.ng) from *N.*

gonorrhoeae:

5	orf84.pep	MAEICLTGTGPGSKTLKQVSMNDEMFKPDEKAIKRVFTNIGLKIPIHTYIETDAKK	60
	orf84.ng	MAEICLTGTGPGSKTLKQVSMNDEMFKPDENGVRKRVFTNIGLKIPIHTYIETDAKK	60
10	orf84.pep	LPKSTDEQLSAHDMYEWIKKPNVIGSIVIVDEAQDVPVARSAGSKIPENVQWLNTHRHG	120
	orf84.ng	LPKSTDEQLSAHDMYEWIKKPNVIGSIVIVDEAQDVPVARSAGSKIPENVQWLNTHRHG	120
15	orf84.pep	IDIFVLVTGPKLLDQNLRLTVRKHYHIAANKMGRMTLLEWKICADDFVKMASSAFSSIYT	180
	orf84.ng	IDIFVLVTGPKLLDQNLRLTVRKHYHIAANKMGRMTLLEWKICADDFVKMASSAFSSIYT	180
20	orf84.pep	LDKKVYDLYXXAEVHTVNVKRSKWFTYLPVIVLLIPFVGLSYKMLSSYGKKQEEPAQ	240
	orf84.ng	LDKKVYDLYSABHTVNVKRSKWFTYLPVIVLLIPFVGLSYKMLSSYGKKQEEPAQ	240
25	orf84.pep	ESAAEQQAVLPDKTEGEPVNNGLTADMVPTLSSEKPKSPIYNGVRQVTFEYIAGCI	300
	orf84.ng	ESAAEQQAVLPDKTEGESVNNGLTADMVPTLPKPKSPIYNGVRQVTFEYIAGCI	300
30	orf84.pep	EGGRTGCACTYSHQGTALKEVTELMCKDYVKNGLFPNPFYKESQGEVQQAQQHSDRAQV	360
	orf84.ng	EGGRTGCTCYSHQGTALKEVTELMCKDYVKNGLFPNPFYKESQGEVQQAQQHSDRAQV	360
35	orf84.pep	ATLGCKPKQNLMYDNWEERKPFEGIGGGVVGSA	395
	orf84.ng	ATLGCKPQNLMYDNWEERKPFEGIGGGVVGSA	395

The complete length ORF84ng nucleotide sequence <SEQ ID 325> is:

1	ATGCGAGAAA	TCTGTTTGAT	AACCGGCAGC	CCCGGTTGAC	GGAAAAACATT
51	AAAAATGGTT	TCCATGATGC	CAACACGATG	AATGTTTAA	CCAGATGAAA
101	ACGGCGTAGC	CGTAAAGCTA	TTTACGACAA	TCAAAGGTTT	GAAGATACCG
151	CACACCCACA	TAGAAACAGA	CGCAAGAAG	CTGCCGAAAT	CAACCGATGA
201	ACAGCTTTTC	GCGGATGATA	TGTATGAATG	GATCAAGAAG	CCTGAAAacg
251	tccggcgCAAT	CGTTATTGTC	GATGAGCGCG	AAGACGTATG	GCCCGCAGCG
301	TccgCAGGTT	CGAAAATCC	CGAAAAAGTC	CAATGGCTGA	ACACACACAG
351	GCATCAGGCG	ATAGATATAT	TTGATTTAGC	ACAAGGTCCT	AAACTCTTAG
401	ATCAGAACTT	CGGACATGTC	GTTTAAAGAC	ATTACACAT	TGCGCGCAAC
451	AAATGGGTT	TGCGTACCTT	GTTTGAATGG	AAGATGATGG	CGGATGACCC
501	GGTAAAAATG	GCATCAAGTG	CATTTCACAG	TATCTACACA	CTGGATAAAA
551	AGTTTATGTA	CTTGACGAAA	TCCGACGAAA	TTACACCGGT	AAACAAAGCT
601	AGCGGTTCAA	AATGGTTTGA	TGCAATGCCC	GTCATCATAT	TATTGATTC
651	GCTATTGTC	GTTTGTCTCT	ACAAAATGTT	GGCGAGTATG	GGAAAAAAC
701	AGGAAGAACC	CCGACACAAA	GAATCGCGCG	CAACAGACAA	CGACGCAGTA
751	CTTCCGATAA	AACAGCAAGG	AGAATCGGTG	AATAACGGAA	ACCTTACGGC
801	AGATATGTTT	GTTCCGACAT	TGCCCGAAAA	ACCGCAAGC	AAGCGGATTT
851	ATAACGGTGT	AAGGCAGGTA	AGGACCTTTG	AATATATAGC	AGGCTGTATA
901	GAAGCGCGAA	GAACCGGATG	CACCTGCTAT	TGCGATCAAG	GGACGGCATT
951	GAAAGAAGTG	ACGGAGTTGA	TGTGCAAGGA	CTATGTAAAA	AACGCGTTGC
1001	CGTTTAAACC	ATACAAAGAA	GAAAGCCCAAG	GGCAGGAAGT	TCAGCAAAAG
1051	GCGCAGCAAC	ATTCCGACAG	GGCGCAAGTT	GCCACCTTGG	GGGAAAAAC
1101	CAGCAGCAAC	CTAATGTACG	ACAAATGGGA	AGAACGGGGG	AAACCGTTTG
1151	AAGGAATCGG	CGGGGGCGTG	GTGCGATCGG	CAAACTGA	

This encodes a protein having amino acid sequence <SEQ ID 326>:

1	MAEICLTGTT	PGSKTLKQV	SMNDEMFK	PDENGVRKV	FTNIGLKIPI
51	HTHIETDAKK	LEKSTDEQLS	AHDYEWIKK	PENGVARIV	DEAQDVPV
101	SAGSKIPENV	QWLNTHRHG	IDIFVLVTG	KLLDONLRL	VKRHYHIAAN
151	KMLRLTLLW	KVCADDVPKM	ASSAFSSIYT	LDKKVYDLYE	SAEHTVNVKV
201	KRSKWFTYLP	VIIILLPLFV	GLSYKMLSGY	GKKQEEPAQ	ESAAEQQAV
251	LPDKTEGESV	NNGNLTADM	VPTLPKPKES	KPIYNGVRQV	TFEYIAGCI
301	EGGRTGCTCY	SHQGTALKEV	TELMCKDYVK	NGLFPNPFYK	ESQGEVQQS
351	AQQHSDRAQV	ATLGCKPQNN	LYDNWEERG	KPFEGIGGGV	VGSAN*

ORF84ng and ORF84-1 show 95.4% identity in 395 aa overlap:

		10	20	30	40	50	60
	orf84-1.pep	MAEICILTCTPGSGKTLKMWMMANDEMFKPDENGIRKRVFTNIKGLKIPHTYIETDAKK					
5	orf84ng	MAEICILTGTGPGSGKTLKMWMMANDEMFKPDENGVRKRVFTNIKGLKIPHTIETDAKK					
		10	20	30	40	50	60
	orf84-1.pep	LPKSTDEQLSAHDMYEWIKKPNIGSIVIVDEAQDVMPARSAGSKIENPVOWLNRHOG					
10	orf84ng	LPKSTDEQLSAHDMYEWIKKPNVGAIVIVDEAQDVMPARSAGSKIENPVOWLNRHOG					
		70	80	90	100	110	120
	orf84-1.pep	LDKVVVDLYESAEVHTVNKVKRSKWFTLPVIVLLIPVTVGLSYMLSSYGKKOEPAQAQ					
	orf84ng	LDKVVVDLYESAEIHTVNKVKRSKWFTLPVIVLLIPVTVGLSYMLSSYGKKOEPAQAQ					
		130	140	150	160	170	180
15	orf84-1.pep	IDIFVLTQGEKLLDQNLRLTLVKRHYHIAANKMGLRLTLKWKVCADDPVMMASAFSSIT					
	orf84ng	IDIFVLTQGEKLLDQNLRLTLVKRHYHIAANKMGLRLTLKWKVCADDPVMMASAFSSIT					
		130	140	150	160	170	180
20	orf84-1.pep	LDKVVVDLYESAEVHTVNKVKRSKWFTLPVIVLLIPVTVGLSYMLSSYGKKOEPAQAQ					
	orf84ng	LDKVVVDLYESAEIHTVNKVKRSKWFTLPVIVLLIPVTVGLSYMLSSYGKKOEPAQAQ					
		190	200	210	220	230	240
25	orf84-1.pep	ESAAETEQAQVLPDKTEGSPVNNGNLTADMFPVPTLSEKPEKPIYNGVRQVRTFEYIAGCI					
	orf84ng	ESAAETEQAQVLPDKTEGSPVNNGNLTADMFPVPTLSEKPEKPIYNGVRQVRTFEYIAGCI					
		250	260	270	280	290	300
30	orf84-1.pep	EGGRGTGCACYSHQGTALKEVTELMCKDVKVNGLPFNPKYKEESQGGVEQVQSAQQHSDRAQV					
	orf84ng	EGGRGTGCTCYSHQGTALKEVTELMCKDVKVNGLPFNPKYKEESQGGVEQVQSAQQHSDRAQV					
		310	320	330	340	350	360
35	orf84-1.pep	ATLGGAPXONLMYDNWEERKPFEGIGGGVVGSA					
	orf84ng	ATLGGAPXONLMYDNWEERKPFEGIGGGVVGSA					
		370	380	390			
40	orf84-1.pep	ATLGGAPXONLMYDNWEERKPFEGIGGGVVGSA					
	orf84ng	ATLGGAPXONLMYDNWEERKPFEGIGGGVVGSA					
		370	380	390			

Based on this analysis, including the presence of a putative transmembrane domain (single-underlined) in the gonococcal protein, and a putative ATP/GTP-binding site motif A (P-loop, double-underlined), it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 39

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 327>:

50	1	GTGGTTTTC	TGAATGCCGA	CAACGGGATA	TTGGTTTCAG	ACTTGCCTTT
	51	TGAAGTCAAA	CTGAAAAAAT	TCATATCGA	TTTTTAACAAT	ACGGGTATGC
	101	CGCGTGATT	CGCCAGCGAT	ATTGAAGTCA	CGGACACGC	AACCGTGAG
	151	AAACTCGAGC	GCACCATCCG	CGTGAACCAT	CCTTTGACCT	TGCACGGCAT
	201	CACGATTTAT	CAGGCGAGTT	TTGCCGACGG	CGGTTGCGAT	TTGACATTCA
	251	AGGCGTGGAA	TTTGGGTGAT	GCTTCGCGCG	AGCCTGTGCT	GTGGAAGCA
55	301	ACATCCATAC	ACGAGTTTCG	GTTGGAATAT	GGCAACACCA	AATATCGTCT
	351	TGAGTTCGAT	CAGTTCACCT	CTTGAATAT	GGAGACATG	ACCGAGCGG
	401	CGGAACCGGA	AAAAGCGCTG	AAATCCAGCG	TGCCCGATGT	CCGCGCGGTT
	451	ACTCAGGAAG	GTCACAAATA	CACCAAT...	TACCG
	501	TATCCGTGAT	GCGCCAGGCC	AGGCGGTGCA	ATATAAAAAC	TATATGCTGC
60	551	CGGTTTTCGA	GGAACAGGAT	TATTTTTCGA	TTACCGGCAC	GCGCAGCGC.

-219-

5 601 TTGCAGCAGC AATACCGCTG GCTGCGTATC CCCTTGGACA AGCAGTTGAA
651 AGCGGACACC TTTATGGCAT TGCCTGAGTT TTTGAAAGAT GGGGAAGGGC
701 GCAACGCTCT . GTTGGCGGAC GCAACCAAAG GCGCACTCGT CGAAATCCGC
751 GAACAATTCA TGCTGGCTCG GGAACACACG CTGACATCTT TTGCACAAAA
801 AGGCTATTGG GATTGGGACG AATTATTATC GTCCAAATAT CGGAAAGAGC
851 AGCAGGATAA GATGCAAGGC TATTCTTACG AAATGCTTTA CGGCGTGATG
901 AACGCTGCTT TGGATGAAAC CAT . ACCCGG TACGGCTTCG CGGAATGGCA
951 GCAGGATGAA GCGGGGAATC GTTTCCTGCT GCACAGTATG GATGCGTACA
1001 CGGGTTTGAC CGAATATCCC GCGCCTATCG TGCTGCACTT TGATGGGTTT
10 1051 TCCGAGTGGC TGTGCTCGGG TTTCGAGATG ACCCGTCCC C . GGTCCGCT
1101 TTTGGTCTAT CTC...

This corresponds to the amino acid sequence <SEQ ID 328; ORF88>:

1 MVFLNADNGI LVQDLPEFVK LKKFHIDFYN TGMPRDFASD IEVTDKATGE
51 KLERTIRVNH PLTLHGITYI QASFADGGSD LTFKAWNLDG ASREPVLKA
101 TSIHQFPLEI GKHKYRLEFD QFTSMNVMD SEGAEREKSL KSTLPDVRV
151 TQEGHKYTNX XXXXYRIRD APQDAVEYN YMLPVLQEQD YFWITGTRX
201 LQQQYRWLRI PLDKQLKADT FMALREFLKD GEGRRKRVAD ATKGAPAEIR
251 EQFMLAENT LNIFAQKGYL GLOBEFTSNI PKEQDQKMQG YFEMLYGVM
301 NAALDETXTR YGLEFWQOQE ARNRFLHSM DAYTGLTETP APMLQLDQGE
351 SEVRSSGLQM TRSKXGPLLV L...

Further work revealed the complete nucleotide sequence <SEQ ID 329>:

1 ATGAGTAAAT CCGCTAGATC TCGCCCACTT GTTCCCGCTG CGTGGTTCCG
51 TTTTTCAGC TCCATGCGCT TTGCACTGCG TTTGCTCAGT CTGCTGGGTA
101 TTGCATCGGT TATCGGTACG GTGCTGCGAG AAAACACGCG CGACAGCGAT
151 TATTTTGGTCA AATTCGGATC GTTTTGGGCG CAGATTTTGT GTTTTCTGGG
201 ACTGTATGAC GTCTATGCTT CGGCTATGTT TGCTGTTATC ATGATGTTTT
251 TGGTGGTTTC TACCACTTTG TGCTGATTC GCAATGTGCC CGCGTCTGCG
301 CGCGAAATGA AGTCTTTTGC GGAAGAGGTT AAGAAATAAT CCTTGGCGGC
351 GATGCGGCAT TCTTCGCTGT TGGATGTAAA AATTGCGCGC GAGGTTGCCA
30 401 AACGTTATCT GGAAGTACAA GGTTTTCAGG GAAAACCAT TAACCGTGAA
451 GACGGGTCCG TTCTGATTGC CGCAGAAAAA GGCACAAATG ACAAATGGGG
501 CTATATCTTT GCCCATGTGT CTTTGATTGT CATTTGCCGT GCGCGGTTGA
551 TAGACAGTAA CCGCTGTTGT AAACCTGGTA TCGTAGCCGC TCGGATTTGT
601 CGCGACAATC AGCGGTTTTA TGCCAAAGAT TCAACCGCGT AAGATATTTT
35 651 GGGTGGCTCC AATCTCTCAT TTAGGGCAA CCGCATATT TCGGAGGGCG
701 AGAGTGCAGG TGTGTTTTC CTGAATGCGC ACAACGGGAT ATTGTTTCAG
751 GACTTGCCCT TTGAAGTCAA ACTGAAAAAA TTCCATATCG ATTTTACAA
801 TACGGGTATG CCGCTGATT TCGCCACGCA TATTGAAGTG ACGGACAAGG
851 CACCGGTGTA GAACTCTGAG CGCACCATCC GCGTGAACCA TCTTTTACCC
901 TTGCACGGCA TCACGATTTA TCAGGCGAGT TTTGCGCAGC GCGGTTGGGA
951 TTTGACATTC AAGCGGTGGA ATTTGGGTGA TGCTTCGCGC GAGCCTGTGC
1001 TGTGTAAGGC AACATCATTA CACCAAGTTT CGTTGGAAAT TGGCAAAACAC
1051 AAATATCGTC TTGAGTTCGA TCAGTTCAC TCTATGAATG TGGAGGACAT
1101 GAGCGAGGCG GCGGAACCGG AAAAAAGCCT GAAATCCACG CTGAACGATG
45 1151 TCCGCGCCGT TACTCAGGAA GGTAAAAAAT ACACCAATAT CGGCCCTTCC
1201 ATTGTTTACC GTATCCGTTG TCGCGCAGGG CAGGCGGTGC AATATAAAAA
1251 CTATATGCTG CCGGTTTTGC AGGCAAGAGA TATATTTTGG ATTACCGGCA
1301 CGCGAGCGAG CTTCAGACAG CAATACCGCT GCGTGTGATC CCCCTTGAAC
1351 AAGCAGTTGA AAGCGACAC CTTTATGCGA TTTGCTGAGT TTTTGAAGAG
50 1401 TGGGGAAGGG CGCAAAACGC TGGTTGCCGA CGCAACCAA GCGCGACCTG
1451 CCGAAATCCG GCAACAATT ATGCTGGCTG CGGAAACAC GCTGACACAT
1501 TTTGCACAAA AAGGCTATT GGGATGGAC GAATTTATTA CGTCCAATAT
1551 CCGGAAGAG CAGCAGGATA AGATGCAAGG CTAATTTCTC GAATGCTT
1601 ACGGCTGAT GAACGCTGCT TTGGATGAAA CCATACGCGC GTACGCGTGT
55 1651 CCGGAATGGC AGCAGGATGA AGCGCGGAAT CGTTTCTCTG TGACACGAT
1701 GGAATGCTAC ACGGTTTTGA CGGAATATCC CGCGCTATG CTGCTGCAAC
1751 TTGATGGGTT TTTCCAGSTG CGTTCGTCGG GTTTGCAGAT GACCGGTCTC
1801 CCGGTCGCGC TTTTGTCTA TCTCGGCTCG GTGCTGTTGG TATTGGGATG
1851 GGTATTGATG TTTTATGTGC GCGAAAAACG GCGGTGGGTA TTGTTTTCAG
60 1901 ACGCGAAAT CCGTTTTGCC ATGTTCTTGC CCGCGAGTGA ACGGATTTG
1951 CAGAGGAAT TTCCAAACCA CGTCCAGAGT CTGCACCGC TCGGCAAGGA
2001 CTGAATCAT GACTGA

This corresponds to the amino acid sequence <SEQ ID 330; ORF88-1>:

5 1 MSKSRSPFL LSRFWFAFFS SMRFAVALLS LLGIASVIGT VLQONQPTD
51 YLVKFGSFNA QLVFGLGLYD VYASAFVVTI MMFLVVTSL CLIRNVFFW

101	REMKSFREKV	KEKSLAAMRH	SSILDVKIAP	EVAKRYLEVO	GFOGKTINRE
151	DGSLVLAARK	GTMMKWGYIF	AHVALIVICL	GGIDSNLLIL	KLGMITGRIV
201	PNDQAVYAKD	FKPESILGAS	NLSFRGNVNI	SEGQSDVVF	INADNGILVQ
251	DLPFVEVLKK	FHIDFYNTGM	PROFASDIEV	TDKATGEKLE	RTIRVNHPLT
301	LHGITIYQAS	FADGSSDITF	KAWNIGDASR	EPVLIKATSI	HQFPLEIGKH
351	KYRLFDQDT	SMNVDMSEG	AEREKSLKST	INDVRAVTEQ	GKKYTNIGPS
401	IVYRIRDAAG	QAVEYKNYML	PVLQEQDYFW	ITGTRSGLQQ	QYRWLRIPLD
451	KQLKADTFMA	LREFLKDGGG	RKRIVADATK	GAPAEIREQF	MLAAETITNI
501	FAQGYLGLD	EFITFNIKE	QQQRMQGYFY	EMLYGVNMNA	LDEITIRYGL
551	FEWQODEARN	RPLHSMOAY	VGLTEYFAFM	LQLQDGSFV	RSSGLQNTFS
601	PGALLVYLG	VLLVLGTVM	FYVREKRAV	LPSDGKIRPA	MSSARSERDL
651	QKEFPKHVES	LQRLGKDLNH	D*		

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

- 15 ORF88 shows 95.7% identity over a 371aa overlap with an ORF (ORF88a) from strain A of *N. meningitidis*:

	orf88.pep				10	20	30
					MVF	INADNGILVQDL	PFEVKLKKFHIDFY
20	orf88a	AKDKFPE	SILGASNLS	FRGNVNI	SEGQSDVVF	INADNGILVQDL	PFEVKLKKFHIDFY
		210	220	230	240	250	260
	orf88.pep		40	50	60	70	80
25	orf88a	TGMRPF	ASDIEVTD	KATGEKLE	RTIRVNHPL	LHGITIYQAS	FADGSSDITF
		270	280	290	300	310	320
	orf88.pep		100	110	120	130	140
30	orf88a	ASREPV	VLIKATSI	HQFPLEIG	KHKYRLFDQ	TSNMVDMSEG	AEREKSLKSTLP
		330	340	350	360	370	380
	orf88.pep		160	170	180	190	200
35	orf88a	TOEGHK	YTNXXXXX	YRIRDA	FGQAVEYKNYML	PVLQEQDYFW	ITGTRSGLQQQYRWLR
		390	400	410	420	430	440
	orf88.pep		220	230	240	250	260
40	orf88a	PLDKQL	KADTFMAL	REFLKDGE	GKRKRVADAT	KGAPAEIREQ	FMLAENTINIF
		450	460	470	480	490	500
	orf88.pep		280	290	300	310	320
45	orf88a	GLDEFIT	SNIPKEQ	QDMQGYFY	EMLYGVNMNA	LDEITIRYGL	PEWQODEARN
		510	520	530	540	550	560
	orf88.pep		340	350	360	370	
50	orf88a	DAYTGL	TEYFAFM	LQLDGFSE	VRSGLQMT	RSXGFLVLYL	
		570	580	590	600	610	620
	orf88a	ANVL	FSDGKIR	FAMSSARS	SERDLQKE	FPKHVESLQ	RGLGKDLNHDX
60		630	640	650	660	670	

The complete length ORF88a nucleotide sequence <SEQ ID 331> is:

```

1  ATGAGTAAAT  CCGFAGATC  TCCCCACTT  CTTTCCCGC  CGTGGTTCGC
51  TTTTTCGAGC  TCAATCGCT  TTGCGTCGC  TTTGCTCAGT  CTGCTGGGTA
101  TTGCATCGT  TATCGGTACG  GTGTGCAGC  AAAACGACCC  GCAGACGGAT

```

5	151	TATTTGSTCA	AATTCGGATC	GTTTTGGGCG	CAGATTTTTC	GTTTCTGGG
	201	ACTGTATGAC	GTCTATGCTT	CGGCATGGTT	TGCTGTATTC	ATGATGTTTT
	251	TGGTGGTTTC	TACCAAGTTG	TGCTGATTC	GCAATGTGCC	GCOSGTTCTGG
	301	CGCGAATATGA	AGTCTTTTCG	GGAAAGAGTT	AAAGAAAAAT	CTCTGCGGCG
	351	GATGCGCCAT	TCTTCGCTGT	TGGATGTAAA	AATTGCGCCC	GAGSTTGCGA
	401	AACGTTATCT	GGAAAGTCAA	GTTTTTCAGG	GAATAACCAT	TAAOCSGTGAA
	451	GACGGGTTCGG	TTCCTGATTG	CGCCAAAAAA	GGCAATGA	ACAAATGGGG
	501	CTATATCTTT	GCCCATGTTG	CTTTGATGTT	CATTGCGCTG	GGCGGGTTGA
	551	TAGACAGTAA	CGTCTGTGTT	GAAGTGGGTA	TGCTGACGCT	TGCGAATGTTT
10	601	CGGACATATC	AGCGGTATTA	TGCCAAGAGT	TTCAAGCCCG	AAAGATTTTT
	651	GGGTGGCTCC	AATCTCTCAT	TTAGGGGCAA	CGTCAATATT	TCGAGGGGCG
	701	AGAGTGGCGA	TGTGTTTTC	CGCAATCGCG	ACAACGGGAT	ATTGTTTCAG
	751	GACTTGCCTT	TTGAAGTCAA	ACTGAAAAAA	TTCCATATCG	ATTTTTTCAA
	801	TACGGGTATG	CGCGCGGATT	TGCGCAGTGA	TATTGAAGTA	ACGGATAAAG
15	851	CAACCGGTGA	GAAACTCGAG	CGCACCATCC	CGGTGAACCA	TCCTTTGACC
	901	TTGCACGGCA	TACAGATTGA	TCAGGCGAGT	TTTGCGGACG	CGGGTTGCGA
	951	TTTGACATTC	AAGGCGTGGG	ATTGTTGGTG	TGCTTGCAGC	GAGCCTGTGC
	1001	TGTTGAAGGC	AACATCCATA	CACCGATTTT	CGTTGGAAAT	TGGCAACAC
20	1051	AAATATCGTC	TTGAGTTCGA	TCAGTTTACT	TCTATGAATG	TGGAGGACAT
	1101	GAGCGAGGGC	GCGGAACGGG	AAAAAAGCCT	GAAATCCACG	CTGAAGCATG
	1151	TCGCGCGCGT	TACTCAGGAA	GGTAAAAAAT	ACACCAATAT	CGGCCCTTCC
	1201	ATTGTTTACC	GTATCTCGTG	TGCGCGAGCG	CGCGGCTCG	AATATAAACA
	1251	CTATATCGTG	CGGTTTTTCG	AGGAACAGGA	TATTTTTTGG	ATTACCGCGA
	1301	CGCGCAGCGG	CTTGACGACG	CAATACGCTG	GGCTGGCATAT	CCCTCTGGAC
25	1351	AAGCAGTTGA	AAGCGACACG	CTTTATGGCA	TTGCGTGAGT	TTTTGAAGA
	1401	TGGGGAGAGG	CGCAACAGTC	TGGTTGCGGA	CGCAACCAA	GGCGCACTTG
	1451	CGGAATTCG	CGCAACATTC	ATGCTGGCTG	CGGAACAAC	GCTGAACATC
	1501	TTTGACAAA	AAGGCTATTT	GGGATGGAC	GAATTTATTA	CGTCCAATAT
	1551	CCGGAAGAG	CAGCAGGATA	AGATGCAGGG	CTATTTCTAT	GAAATGCTTT
30	1601	ACGCGGTGAT	GAAACGCTGCT	TTGGATGAAA	CCATACCGCG	GTACGGCTTG
	1651	CCGGAATGGC	AGCAGGATGA	AGCGCGGAAT	CGTTTCTCGT	TGCACAGTAT
	1701	GGATGCGTAC	ACGGGTTTTGA	CGGATATCC	CGCGCCTATG	CTGCTGCAAC
	1751	TTGATGGGTT	TTCCGAGGTG	CGTTCTGCGG	GTTTGAGAT	GACCCGTTCC
	1801	CGGGTGGG	TTTTGCTCTA	TCTCGGCTCG	TGCTGTTG	TATTTGGTAT
35	1851	GGTATTGATG	TTATCTCGTG	CGGAACAAC	GGCTGGGTA	TGTTTTGAG
	1901	ACGGGAATAT	CGGTTTTTCG	ATGCTTTCG	CCGCAAGCA	ACGGGATTTG
	1951	CAGAAGGAT	TTCCAAACAA	CGTCGAGAGT	CTGCAACGCG	TGGCAAGGA
	2001	CTTGAATCAT	GACTGA			

This encodes a protein having amino acid sequence <SEQ ID 332>:

40	1	MSKRRSPPL	LSRPWFAFFS	SMRFAVALLS	LLGIASVIGT	VLOQNQPTOD
	51	YLWKFSGFWA	QIFGFLGLGYO	VIASAWFVVI	MMFLVSTSL	CLIRNVPPFW
	101	REMKSFREKV	KEKSLAAMRH	SLLLDVKIAP	EVAKRYLEVQ	GFQKGTINRE
	151	DGSLVIAAKK	GTMNKNGYIF	AHVALIVICL	GGLIDSNLL	KLGLMTGRIV
	201	PDNQAVYAKD	FKPESILGAS	NLSFRGNVNI	SEGOSADVVF	LNADGILVQ
45	251	DLFPFVKLLK	FHIDFYNTGM	PRDFASDIEV	TOKATGEKLE	RTIRVNHPLT
	301	LHGHTIYQAS	FADGGSDLTFF	KAWNLDGASR	EPVVLKATSI	HQFPLEIGKH
	351	KYLEPFDQIT	SMNVEDMSEG	AEEREKSLKST	LNDVRATVGE	GKRYTHIGPS
	401	IVVIRRDAG	QVRYVNVML	PVLQODYFW	ITGTRSGILQ	QYWLRIPLD
	451	KOLKADTFMA	LRFLKDGEG	KRLVADATK	GAPADIREOF	MLAENTINI
50	501	PAQKGYLGLD	EFITSNIPKE	QODKMQGYFY	EMLYGVMNAA	LDETIRRYGL
	551	PEWQOQEARN	RFLHLSMDAY	TGLTEYPAM	LQLDGFSEV	RSSGLQMYRS
	601	PGALLVYLGS	VLLVLGTVLM	FVYREKRAWV	LFSDGKIRFA	MSSARSERDL
	651	QKEFFKHVES	LQRLGKDLNH	D*		

ORF88a and ORF88-1 100.0% identity in 671 aa overlap:

55	orf88a.pep	MSKRRSPPLLSRPWFAFFSSMRFAVALLSLLGIASVIGTVLQONQOPQTDYLVKFGSFWA	60
	orf88-1	MSKRRSPPLLSRPWFAFFSSMRFAVALLSLLGIASVIGTVLQONQOPQTDYLVKFGSFWA	60
60	orf88a.pep	QIFGFLGLGYDVYASAWFVVIIMFLVSTSLCLIRNVPPFWRMKSFREKVKESLAAMRH	120
	orf88-1	QIFGFLGLGYDVYASAWFVVIIMFLVSTSLCLIRNVPPFWRMKSFREKVKESLAAMRH	120
	orf88a.pep	SLLLDVKIAPEVAKRYLEVQGFQKGTINREDGSLVIAAKGTMNKGWYIFAHVALIVICL	180
65	orf88-1	SLLLDVKIAPEVAKRYLEVQGFQKGTINREDGSLVIAAKGTMNKGWYIFAHVALIVICL	180
	orf88a.pep	GGLIDSNLLKLGLMTGRIVPDNQAVYAKDFKPESILGASNLSFRGNVNISEGOSADVVF	240

			240
5	orf88-1	GGLLDSNLLKLGMLTGRIVPDNQAVYAKDFKPESILGASNLISFRGNVNISEGQSDVVF	240
	orf88a.pep	LNADNGILVQDLFFEVLKKKFHIDFYNTGMPDRDFASDIEVTDKATGEKLEIRTVNHPLT	300
	orf88-1	LNADNGILVQDLFFEVLKKKFHIDFYNTGMPDRDFASDIEVTDKATGEKLEIRTVNHPLT	300
	orf88a.pep	LHGITTYQASFADGSDLTFAKWNLGDSAREPVLKATSIHQFPLEIGHKKHYRLEFDQFT	360
10	orf88-1	LHGITTYQASFADGSDLTFAKWNLGDSAREPVLKATSIHQFPLEIGHKKHYRLEFDQFT	360
	orf88a.pep	SMNVEDMSGAEREKSLKSTLNDVRAVTQEGKKYTNIGPSIVYRIRDAAGQAVEYKNYML	420
15	orf88-1	SMNVEDMSGAEREKSLKSTLNDVRAVTQEGKKYTNIGPSIVYRIRDAAGQAVEYKNYML	420
	orf88a.pep	PVLQEQQDYFWITGTRSGLQQQYRWLRIPLDKQLKADTFMALREFLKDGEGRKRLVADATK	480
	orf88-1	PVLQEQQDYFWITGTRSGLQQQYRWLRIPLDKQLKADTFMALREFLKDGEGRKRLVADATK	480
20	orf88a.pep	GAPAEIREQFMMLAAENTLNIFAQKGYLGLDEFITSNIPKEQQQDMQGYFYEMLYGVNNA	540
	orf88-1	GAPAEIREQFMMLAAENTLNIFAQKGYLGLDEFITSNIPKEQQQDMQGYFYEMLYGVNNA	540
25	orf88a.pep	LDETIRRYGLPEWQQDEARNRFLHLSMDAYTGLTEYPAPMLQLDGFSEVRSSGLQMTS	600
	orf88-1	LDETIRRYGLPEWQQDEARNRFLHLSMDAYTGLTEYPAPMLQLDGFSEVRSSGLQMTS	600
	orf88a.pep	PGALLVYLGSVLLVLTGLVLMFYVREKRAWLFSDGKIRFAMSSARSERDLQKEFPKHVES	660
30	orf88-1	PGALLVYLGSVLLVLTGLVLMFYVREKRAWLFSDGKIRFAMSSARSERDLQKEFPKHVES	660
	orf88a.pep	LQRLGKDLNHD 672	
35	orf88-1	LQRLGKDLNHD 672	

Homology with a predicted ORF from *N.gonorrhoeae*

ORF88 shows 93.8% identity over a 371aa overlap with a predicted ORF (ORF88.ng) from *N. gonorrhoeae*:

40	orf88.pep	MVFLNADNGILVQDLFFEVLKKKFHIDFYNTGMPDRDFASDIEVTDKATGEKLEIRTVNH	60
	orf88ng		
	orf88a.pep	PLTLHGITTYQASFADGSDLTFAKWNLGDSAREPVLKATSIHQFPLEIGHKKHYRLEFD	120
45	orf88ng	PLTLHGITTYQASFADGSDLTFAKWNLDASREPVLKATSIHQFPLEIGHKKHYRLEFD	120
	orf88.pep	QFTSMNVEDMSGAEREKSLKSTLPDVRVAVTQEGHKYTNXXXXXXYYRIRDAAGQAVEYKN	180
	orf88ng	QFTSMNVEDMSGAEREKSLKSTLNDVRAVTQEGKKYTNIGPSIVYRIRDAAGQAVEYKN	180
50	orf88.pep	YMLPVLQEQQDYFWITGTRSLQQQYRWLRIPLDKQLKADTFMALREFLKDGEGRKRLVAD	240
	orf88ng		
	orf88a.pep	ATKGAPAEIREQFMMLAAENTLNIFAQKGYLGLDEFITSNIPKEQQQDMQGYFYEMLYGVN	300
55	orf88ng	ATKDAPAEIREQFMMLAAENTLNIFAQKGYLGLDEFITSNIPKEQQQDMQGYFYEMLYGVN	300
60	orf88a.pep	NAALDETXYTRYGLPEWQQDEARNRFLHLSMDAYTGLTEYPAPMLQLDGFSEVRSSGLQ	360
	orf88ng	NAALDETIRRYGLPEWQQDEARNRFLHLSMDAYTGLTEYPAPMLQLDGFSEVRSSGLQ	360
	orf88a.pep	TRSXGFLVYL	371
65	orf88ng	TRSEFALLVYLGSVLLVLTGLVLMFYVPPKRAWLFSNXXIRFAMSSARSERDLQKEFPKH	420

An ORF88ng nucleotide sequence <SEQ ID 333> was predicted to encode a protein having amino acid sequence <SEQ ID 334>:

```

1  MIVFNADNGM IVQDLPEVFK LKKFHIDFYN TGMPRDFASD IEVTDKATGE
5  51 KLRTITRVNH PLTIHGHTIY QASFDAGGSD LTFKAWNLRD ASREPVPVLKA
101 TSHQFPLEI GHKHYRLEFD QFTSMNVEDM SEGAEREKSL KSTINDVRVA
151 TQEGKKYTHI GPSIVYRIRD AAGQAVEYKN YMLPILODKD YFWLTGTRSG
201 LQQQYRWLRI PLDKQLKADT FMALREFLKD GEGRKRLVAD ATKDAFAEIR
251 EQFMLAAGENT INIFAQKGYL GLDEFTISNI PKGQDQKMQG YFEMYELVGM
301 NAALDETIIR YGLPEWQDGE ARNRFLHSM DAYTGLTEYP AFMLQLQDGF
10 351 SEVRSSSLQCM TRSPGALLVY LGSVLLVLGT VFMFYVPPKKR AWVLFNSKKI
401 RFAMSSARSE RDLQKEFFKH VESLQRLGKD LNHD*

```

Further work revealed the complete gonococcal DNA sequence <SEQ ID 335>:

```

1  ATGAGTAAAT CCGCTATATC TCCACACACT TTTCCTCCGC CGTGGTTCGC
15  51 TTTTTCGAGC TCGATGCGCT TTGCGGTCCG TTTGCTCAGT CTGCTGGGTA
101 TTGCATCGGT TATCGGACAG GTGTTACAGC AAAACACAGC SCACAGCGAT
151 TATTGGTGA AATTCGGACC GTTTGGACT CGGATTTTGG ATTTTTCGGG
201 TTTGTATGAT GTCTATGCTT CGGACTGGTT TGTCGTTATC ATGATGTTTC
251 TGGTGGTTTC TACCAGTTTG GTTTTAATCC GUAACGTTC CGCGTTTGGC
301 CGCGAAATGA AGTCTTTCGG GAAAAGGTTT AAAAAGAAAT CTCTGGCGGC
20 351 GATGCGGCAT TCTTCGTGT GTGATGTAAA AATTCGCCCC GAAGTTGCCA
401 AACGTTATCT GGAGCTCGGG GTTTTTCAGG GAAAACCGT CAGCGGTGAG
451 GACGGGTGCG TTCTGATTGC CGCCAAAAAA GGCACaatga acaaATGGGG
501 CTATATCTTT GCccaatag cTTTGATTGT CATTTCGCTG GCGGGTITGA
551 TAGACAGTAA CTTGCTGCTG AAGCTGGGTA TGCTGGCGCG TCGGATTGTT
25 601 COGACAAATC AGGCGGTTTA TGCCAAGATG TTCAAGCCCG AAAGTATTTT
651 GGGTGGCTCC AATCTCTCAT TTAGGGGCAA CGTCAATATT TCCGAGGGGC
701 AAAAGTCGGA TGTCGTTTC CTGAATCGCG ACACCGGAT GTTGGTTCAG
751 GACTTGGCTG ACTGAARAAA TTCCATATCG ATTTTTCAGT
801 TACGGCATTC CGCGCGGATT TTGCCAGGCA TATTGAAGTA ACGCAACAGG
30 851 CAACCGGTGA GAACTCGAG CGCACCATCC CGGTGAACCA TCCTTTGACC
901 TTGCACGGCA TCACGATTTA TCAGGCGAGT TTTCGCGAGC GCGGTTCCGA
951 TTTGACATTC AAGCGGTGGA ATTTGAGGGA TGCTTCGCGC GAACCTGTGC
1001 TGTGTAAGGC AACCTCCATA CACCAAGTTT CGTTGGAATC CGGCAACAC
35 1051 AAATATCGTC TTGAGTTCGA TCAGTTCACT TCTATGAATG TGGAGGACAT
1101 GAGCGAGGGT GCGGAACGGG AAAAAAGCCT GAAATCCACT CTGACGATG
1151 TCCGCGCGCT TACTCAGGAA GGTAAAAAAT ACACCAATAT CGGCGCTTCC
1201 ATCGTGTACC GCATCCGTGA TGcggCAGGG CAGGCGGTGC AATATAAAAA
1251 CTATATGCTG CGATTTTTCG AGGACAAGA TTATTTTGGC CTGACCGGCA
1301 CGCGCAGCGG CTTGCAGCAG CAATACCGCT GGCTGCGTAT CCCCTTGGAC
40 1351 AAGCAGTTGA AAGCGGACAG CTTTATGGCA TTGCGTGAGT TTTTGAAGA
1401 TGGGAAGGCG GCGAAAGTC TGCTTCCGCA CGCACACGTA GACACACGTC
1451 CGGAATCCGC CGAACAATTC ATCTGGCTG CGGAAACAC GCTGAATATC
1501 TTTGCGCAAA AAGGCTATTT GGGATTCGAC GAATTTATTA CTTCCAATAT
1551 CCGGAAGGGC CAGCAGGATA AGATGCAAGG CTATTCTTAC GAAATGCTTT
45 1601 ACGGCGTGAT GAACGCTGCT TTGGATGAAA CCATAGCGCG CTGACGGTTG
1651 CCGGAATGGC AGCAGGATGA AGCGCGGAAC CGTTTCTCGT TGCACAGTAT
1701 GGATGCTATC ACGGGGCTGA CGGAATATCC CGCGCTATG CTGCTCCAGC
1751 TTGACGGGTT TTCCGAGGTC CGTTCCTCAG GTTTCAGAT GACCCGTTGC
1801 CCGGGTGGCG TTTTGGTCTA Tctcggtctg ctattgttgg TTTTGGTctg
50 1851 ggtaTttatg tTTTATGTGC GCGAAAACG GCGGTGGgta tTGTTTTCag
1901 aCGGCAAAAT CCGTTTTGCT ATGCTCTcgg CCcgacgca ACGGGATTTG
1951 aGAAgggaat TTCCAAACCA CGtcgAGAGC CTGCAACggc tcggaagga
2001 CtgaaTCAT GACTga

```

This corresponds to the amino acid sequence <SEQ ID 336; ORF88ng-1>:

```

1  MSKSRIPTL LSRPWEAFSS SMRFVALLS LLGIASVIGT VLQNGQPOTD
55  51 YLVKFPFWT RIFDLGLYD VYASAWFVYI MMELVYSTSL CLIRNVPPFW
101 REMKSFREKV KEKSLAAMRH SLLDVKIAP EVAKRYLEVR GFQKTVSRE
151 DGSVLIAAKK GTMNKMGYIF AQVALIVICL GGLIDSNLLL KLGLMAGRIV
201 PMQAVYAKD FKPEISILGAS NLSFRGNVNI SEGQSDVVF INADNMLVQ
60 251 DLPFEVFLKK FHIDFYNTGM FRDFASDIEV TDKATGEKLE RTIRNVNHLT
301 LHGITIYQAS FADGGSDLTf KAWNLRDASR EPVULKATSI HQFFLEIGHK
351 KYRLEFDQFT SMNVEDMSEG AREKSLKST LNDVRAVTOG GKKYTNIGPS
401 IVYRIRDAAG QAVEYKNYML PILQDKDYFW LTGTRSLQLO QYRWLRPLD
451 KQLKADTFMA LREFLKDGBG RKRLVADATK DAPAZIREQF MLAANTLNI
65 501 FAKGYYGLGD EFTISNPKG QQDKMQGYFY BMLYVNNAA LDETIIRYGL

```

-224-

551 PEWQQDEARN RFLLSHSDAY TGLTEYPAPM LLQLDGFSEV RSSGLQMTRS
601 PGALLVYLGS VLLVLGTVM FYVREKRAWV LFSDGKIRFA MSSARSERDL
651 QKEFPKHVES LQRLGKDLNH D*

ORF88ng-1 and ORF88-1 show 97.0% identity in 671 aa overlap:

5	orf88-1.pep	MSKSRSPPLSRPFWFAFFSSMRFAVALLSLGLASVIGTVLQQNQFQT DYLVKRGSFWA	60
	orf88ng-1	MSKSRISPTLLSRPFWFAFFSSMRFAVALLSLGLASVIGTVLQQNQFQT DYLVKRGSFWT	60
10	orf88-1.pep	QIFGFLGLDYVYASAWFVNVIMFLVYSTSLCLIRNVPFWREMKSFREKVKESLAAMRH	120
	orf88ng-1	RIFDFLGLDYVYASAWFVNVIMFLVYSTSLCLIRNVPFWREMKSFREKVKESLAAMRH	120
	orf88-1.pep	SSLLDVKIAPEAVAKRYLEVQGFGQKTINREDGSVLIAAKKGTMNKWGYFAFHALVIVICL	180
15	orf88ng-1	SSLLDVKIAPEAVAKRYLEVRGFGQKTVSRDGSVLIAAKKGTMNKWGYFAFQALVIVICL	180
	orf88-1.pep	GGILDSNLLLLKGLHGTGRIVPNQVAKAFKPESTLGASNLFRGNVNI SEGQSAOVVF	240
20	orf88ng-1	GGILDSNLLLLKGLHAGRIVPDQNAVYAKDFKPESTLGASNLFRGNVNI SEGQSAOVVF	240
	orf88-1.pep	LNADNGILVQDLFFFEVKLKKFHIDFYNTGMPRDFASDIEVTDKATGEKLESTRVNHPLT	300
	orf88ng-1	LNADNGMLVQDLFFFEVKLKKFHIDFYNTGMPRDFASDIEVTDKATGEKLESTRVNHPLT	300
25	orf88-1.pep	LHGITIVQASFADGGSDLTFKAWNLDASREVPVLKATSIHQFPLEIGHKHYRLFQDFT	360
	orf88ng-1	LHGITIVQASFADGGSDLTFKAWNLDASREVPVLKATSIHQFPLEIGHKHYRLFQDFT	360
30	orf88-1.pep	SMNVEDMSGEAREKSLKSTLNDVRAVTPGKKYTNIGPSIVYRIRDAAGQAVEYKNYML	420
	orf88ng-1	SMNVEDMSGEAREKSLKSTLNDVRAVTPGKKYTNIGPSIVYRIRDAAGQAVEYKNYML	420
	orf88-1.pep	PVLQKQDYFWITGTRSGLQQQYRWLRILPLQKLKATFMALREFLKGSGGKRLVADATK	480
35	orf88ng-1	PILQKQDYFWLTGTRSGLQQQYRWLRILPLQKLKATFMALREFLKGSGGKRLVADATK	480
	orf88-1.pep	GAPAEIREQFMALAEANTLINFAQKGYLGLDEFTSNIPKQGQDMQGYFYEMLYGVMNAA	540
40	orf88ng-1	DAPAEIREQFMALAEANTLINFAQKGYLGLDEFTSNIPKQGQDMQGYFYEMLYGVMNAA	540
	orf88-1.pep	LDETIRRYGLPEWQGDQEARNRFLHSMADYTLGTEYPAPMLQLQDGFSEVRSGLQWTRS	600
	orf88ng-1	LDETIRRYGLPEWQGDQEARNRFLHSMADYTLGTEYPAPMLQLQDGFSEVRSGLQWTRS	600
45	orf88-1.pep	PGALLVYLGSVLLVLTGVIMFVREKRAWLFSDGKIRFAMSSASERDLQKEFPKHVES	660
	orf88ng-1	PGALLVYLGSVLLVLTGVIMFVREKRAWLFSDGKIRFAMSSASERDLQKEFPKHVES	660
50	orf88-1.pep	LQRLGKDLNHD 671	
	orf88ng-1	LQRLGKDLNHD 671	

Furthermore, ORG88ng-1 shows homology with a hypothetical protein from *Aquifex aeolicus*:

gi|2984296 (AE000771) hypothetical protein [Aquifex aeolicus] Length = 537
55 Score = 94.4 bits (2731), Expect = 2e-18
Identities = 91/334 (27%), Positives = 159/334 (47%), Gaps = 59/334 (17%)
Query: 16 FAFSSSMRFVAVALLSLGIASVIG-TVLQNGQPQTDVLKFGPFWTRIFDFLGLVDVYAS 74
+ F S++ + A + ++ +IGT S++G T++QNG YL +FG L D LV+ S
Sbjct: 80 YDPLASLKLAIITMLVLGLLSMIGSTYIRQNQSFWEYLDQFGYDVGIIWKIWLNDVFHS 139
60 Query: 75 AWFVVMIMFLVVSLSCLIRNPPFVREMKSIFREKVKEKSLAARMSSLLDVKIAPAEVAK 134
+ + + + + L V+ + + + + E WIG S + E + + + VRI P +
Sbjct: 140 WYLFIFVLVLNVLNFI SCDSKRLRPFVWQKFS-KERILKLDEAHKILGIPITVKI-PDKR 197
65 Query: 135 -RYLLEVRGFGQGTVFRSGDGLVIAAKGKTMNNKGQIFPAQVALVICGLGIDLSNLLK 192
++L +GF+ V E ++ + +A+KG ++ G ++AL+VI+ G IID
Sbjct: 198 VLKFLFKGKGF-VFVEEKKLVLFAEKGKFSRLGIVYTHIALLVLMAGALID----- 249

Query: 193 GMLAGRIVPONQAVYAKDFKPESILGASNLSEFRGNVNISEGQSADVVFLNADNGMLVQDL 252
 +I+G RG++ ++EG + DV+ + A+ L
 Sbjct: 250 -----AIVGV-----RGLSIVAEGDTNDVNLVGAE--QKPYKL 280

Query: 253 PFEVKLKKFHIDFY---NTGMPEDFA-----SDIEVTDKATGEKLER--TIRVNHPLT 300
 EF V L F I Y N + FA SDIE+ + G K+E T++VN P
 Sbjct: 281 PFAVHLIDFRIKTYAENPNVDKRFQAQVSSYESDIEIIN---GGRVEAKGTVKVNEPFD 337

Query: 301 LHGITIYQASFA---DGGSDLTFKAWNLRDASREP 332
 ++QA++ DG S + + A +P
 Sbjct: 338 FGRYRLFQATYGILDGTSGMGVIVVDKKAHEDP 371

Based on this analysis, including the putative transmembrane domain in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 40

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID 337>:

1 ATGATGAGTA ATAAATGGm AAAAAAGGG TTTACATTGA TTGmGTGAT
 51 GATAGTCTC GCGATACTCG GCATTATCAG CGTCATTGCG ATACCTTCTT
 101 ATCAAAGTTA TATTGAAAAA GGCTATCAGT CCGAGCTTTA TACGAGATG
 151 GTCGGTATCA ACAATATTTC CAACACGTTT ATTTTGAAAA ATCCCTTGGA
 201 CGATAATCAG ACCATCGAGA ACAACTGGA AATATTGTC TCAGGCTATA
 251 AGATGAATCC GAAATTTGCC AAAAAATATA GTGTTTCGGT AAAGTTTGTC
 301 GATAAGGAAA AATCAAGGCC ATACAGTTG GTCCGGCTTC CGAAGCGGG
 351 GACGGGTAT ACTTTGTCCG TATGGATGAA CAGCGTGGC GACGGATACA
 401 AATGCCGTGA TGCCGCTTCT GCCCAAGCCC ATTTGGAGAC CTTGCTCTCA
 451 GATGTCCGCT GTGAAGCCTT CTCTAATCGT AAAAAATAA

This corresponds to the amino acid sequence <SEQ ID 338; ORF89>:

1 MMSNMXXQKG FTLIXMIVV AILGLISVIA IPSYXSIEK GYQSQLYTEM
 51 XGINNISQKF ILKNPLDDNQ TIENKLEIFV SGYKMNPKIA KKYSVSVKVF
 101 DKEKSRAYRL VGVPRAGTGY TLSVWMSVVG DGYKCRDAAS AQAHLETLS
 151 DVGCEAFSNR KK*

Further work revealed the complete nucleotide sequence <SEQ ID 339>:

1 ATGATGAGTA ATAAATGGA AAAAAAGGG TTTACATTGA TTGAGATGAT
 51 GATAGTCTC GCGATACTCG GCATTATCAG CGTCATTGCG ATACCTTCTT
 101 ATCAAAGTTA TATTGAAAAA GGCTATCAGT CCGAGCTTTA TACGAGATG
 151 GTCGGTATCA ACAATATTTC CAACACGTTT ATTTTGAAAA ATCCCTTGGA
 201 CGATAATCAG ACCATCGAGA ACAACTGGA AATATTGTC TCAGGCTATA
 251 AGATGAATCC GAAATTTGCC AAAAAATATA GTGTTTCGGT AAAGTTTGTC
 301 GATAAGGAAA AATCAAGGCC ATACAGTTG GTCCGGCTTC CGAAGCGGG
 351 GACGGGTAT ACTTTGTCCG TATGGATGAA CAGCGTGGC GACGGATACA
 401 AATGCCGTGA TGCCGCTTCT GCCCAAGCCC ATTTGGAGAC CTTGCTCTCA
 451 GATGTCCGCT GTGAAGCCTT CTCTAATCGT AAAAAATAA

This corresponds to the amino acid sequence <SEQ ID 340; ORF89-1>:

1 MMSNMXXQKG FTLEMMIVV AILGLISVIA IPSYXSIEK GYQSQLYTEM
 51 VGINNISQKF ILKNPLDDNQ TIENKLEIFV SGYKMNPKIA KKYSVSVKVF
 101 DKEKSRAYRL VGVPRAGTGY TLSVWMSVVG DGYKCRDAAS AQAHLETLS
 151 DVGCEAFSNR KK*

Computer analysis of this amino acid sequence gave the following results:

Homology with PilE of *N. gonorrhoeae* (accession number Z69260).

ORF89 and PilE protein show 30% aa identity in 120a overlap:

```

orf89 8  QKGFTLIXXMIIVVAILGIISVIAIPSYXSYIEKGYQSQLYTEMXGINNISKQFILKNPL- 66
      QKGFTLI  MIV+AI+GI++ +A+P+Y Y + S+ G + ++ L + +
File 5  QKGFTLIELMIVIAIVGILAAVALFAYQDYTARAQVSEAILLAEGQKSAVTEYLLNHGIW 64

5  orf89 67 --DDNQTIENKLEIFVSGYKMNPKIAKKYSVSVKVDKEKSRAYRLVGVPKAGTYTSLVM 125
      DN + +G + KI KY SV + GV K G LS+W
File 65 FKNDS-----AGVASSDKINGKIVQSVTVAKGVVTAEMASTGVNKEIQGKLSLW 115

```

Homology with a predicted ORF from *N.meningitidis* (strain A)

- 10 ORF89 shows 83.3% identity over a 162aa overlap with an ORF (ORF89a) from strain A of *N. meningitidis*:

```

      10      20      30      40      50      60
orf89.pep  MMSNXXMQKGFLLIXXMIIVVAILGIISVIAIPSYXSYIEKGYQSQLYTEMXGINNISKQF
15 orf89a    MMSNKMKGKGFLLIXXXXXAIXXXSVIXXXSYIEKGYQSQLYTEMXGINNISKQX
      10      20      30      40      50      60
orf89.pep  ILKNPLDDNQTIENKLEIFVSGYKMNPKIAKKYSVSVKVDKEKSRAYRLVGVPKAGTGY
20 orf89a    ILKNPLDDNQTIKSKLEIFVSGYKMNPKIAEKYNVSVHFVNEEKPRAYSLVGVPKTGTGY
      70      80      90      100     110     120
25 orf89.pep  TLSVWMNSVGDGYKCRDAASAQAHALETLSDDVGCFAFSNRKKX
orf89a      TLSVWMNSVGDGYKCRDAASARAHALETLSDDVGCFAFSNRKKX
      130     140     150     160

```

The complete length ORF89a nucleotide sequence <SEQ ID 341> is:

```

30 1 ATGATGAGTA ATAAATGGA AAAAAAGGG TTTACATGTA TTGNGANGNT
51 NATNGNCNTC GCGATACNCN GCNTTANCAG CGTCATTNCN ATNNNTNCNT
101 ATCNGAGTTA TATTGAAAAA GCGTATCACT OCCAGCTTTA TACGGAGATG
151 GTCGGTATCA ACAATATTTT CAAACAGTNT ATTTTGTAAA ATCCCTCGGA
201 CGATAATCAG ACCATCAAGA GAAAACCTGA AATATTTGTC TCAGCGGTATA
35 251 AGATGAATCC GAAAATTTGCC GAAAATATA ATGTTTCGCT GCATTTTGTC
301 AATGAGGAAA AACCNAGGCC ATACAGCTTG GTCGGCGTTC CAAAGACGGG
351 GAOCGGTTAT ACTTTGTCGG TATGATGATA CAGCGTGGGC GACGGATACA
401 AATGCCGTGA TGCCGCTTCT GCCCGAGCCC ATTTGAGAC CTTGCTCTCA
451 GATGTCGGCT GTGAAGCCTT CTCTAATCGT AAAAAATAG

```

- 40 This encodes a protein having amino acid sequence <SEQ ID 342>:

```

1 MMSNKMKGKGFLLIXXXXXAIXXXSVIXXXSYIEKGYQSQLYTEM
51 VGINNISKQX ILKNPLDDNQ TIKSKLEIFV SGYKMNPKIA EKYNVSVHFV
101 NEEKPRAYSL VGVPKTGTGY TSLVWMNSVG DGYKCRDAAS ARAHALETLS
151 DVGCEAFS NR KK*

```

- 45 ORF89a and ORF89-1 show 83.3% identity in 162 aa overlap:

```

      10      20      30      40      50      60
orf89a.pep  MMSNKMKGKGFLLIXXXXXAIXXXSVIXXXSYIEKGYQSQLYTEMXGINNISKQX
50 orf89-1    MMSNKMKGKGFLLIEMMIVVAILGIISVIAIPSYQSYIEKGYQSQLYTEMXGINNISKQF
      10      20      30      40      50      60
orf89a.pep  ILKNPLDDNQTIKSKLEIFVSGYKMNPKIAEKYNVSVHFVNEEKPRAYSLVGVPKTGTGY
55 orf89-1    ILKNPLDDNQTIENKLEIFVSGYKMNPKIAKKYSVSVKVDKEKSRAYRLVGVPKAGTGY
      70      80      90      100     110     120
60 orf89a.pep  TLSVWMNSVGDGYKCRDAASARAHALETLSDDVGCFAFSNRKKX
orf89-1      TLSVWMNSVGDGYKCRDAASAQAHALETLSDDVGCFAFSNRKKX

```


130 140 150 160

Homology with a predicted ORF from *N.gonorrhoeae*ORF89 shows 84.6% identity over a 162aa overlap with a predicted ORF (ORF89.ng) from *N.*5 *gonorrhoeae*:

```

orf89      MMSNMXXQKGFLLIXXMIIVVAILGIISVIAIPSYQSYIEKGYQSQLYTEMVGINNISKQF 60
orf89ng    MMSNMMEQKGFLLIEMMIIVVTLGIISVIAIPSYQSYIEKGYQSQLYTEMVGINNVLKQF 60
10 orf89      ILKNPLDDNQTIENKLEIFVSGYKMNPKIAKKYSVSVKFDVEKESRAYRLVGVPKAGTGY 120
orf89ng    ILKNPQDDNDTLKSKLKI FVSGYKMNPKIAKKYSVSVRFVDAEKPRAYRLVGVPNACTGY 120
15 orf89      TLSVWMNSVCDGYKCRDAASAQAHLTSSDVGCFAFSNRKK 162
orf89ng    TLSVWMNSVCDGYKCRDATSAQAYSDTLSADSGCEAFSNNRKK 162

```

The complete length ORF89ng nucleotide sequence <SEQ ID 343> is:

```

1  atGATGAGCA ATAAATGGA ACAAAGGG TTTACATTGA TTGAGATGAT
51  GATACCTGTC AGCACTACG GCACTATCC CCGTATGCC ATACCTTCTT
20 101 ATCAGACTTA TATTGAAAAA GCCTATCAGT CCCAGCTTTA TACCGAGATG
151 GTCCGTATCA ACAATCTCTT CAAACAGTTT ATTTTGAAA ATCCCCAGGA
201 CGATATGAT ACCCTCAAGA GCAAACTGAA AATATTTCCT TCAGCGTATA
251 AGATGAATCC GAAATattGCC AAAAAATATA GTCTTTCGGT aaggtttGTC
301 gatCGCGAAA AACCAAGGGC ATACGAGTTG GTCCGCGTTC CGAACGCGGG
25 351 GACCGGTTAT ACTTTGTCGG TATGGATGAA CAGCGTGGC GACGATACA
401 AATCGCGTGA TGCCACTTCT GCCCGAGCCT ATTCGACAC CTTGTCGCA
451 GATAGCGGCT GTGAAGCTTT CTCTAATCGT AAAAAATAG

```

This encodes a protein having amino acid sequence <SEQ ID 344>:

```

1  MMSNMMEQKG FTLIEMMIIV TILGIISVIA IPSYQSYIEK GYQSQLYTEM
30 51  VGINNVLKQF ILKNPQDDND TLKSKLKIFV SGYKMNPKIA KKYSVSVRFV
101 DAEKPRAYRL VGVENACTGY TLSVWMNSVG DGYKCRDATS AQAYSDTLSA
151 DSGCEAFSNNR KK*

```

This gonococcal protein has a putative leader peptide (underlined) and N-terminal methylation site (NMePhe or type-4 pili, double-underlined). In addition, ORF89ng and ORF89-1 show 88.3%

35 identity in 162 aa overlap:

```

10 20 30 40 50 60
orf89-1.pep MMSNMMEQKGFLLIEMMIIVVAILGIISVIAIPSYQSYIEKGYQSQLYTEMVGINNISKQF
orf89ng      MMSNMMEQKGFLLIEMMIIVVTLGIISVIAIPSYQSYIEKGYQSQLYTEMVGINNVLKQF
40 10 20 30 40 50 60
orf89-1.pep ILKNPLDDNQTIENKLEIFVSGYKMNPKIAKKYSVSVKFDVEKESRAYRLVGVPKAGTGY
45 orf89ng    ILKNPQDDNDTLKSKLKI FVSGYKMNPKIAKKYSVSVRFVDAEKPRAYRLVGVPNACTGY
70 80 90 100 110 120
orf89-1.pep TLSVWMNSVCDGYKCRDAASAQAHLTSSDVGCFAFSNRKKX
50 orf89ng    TLSVWMNSVCDGYKCRDATSAQAYSDTLSADSGCEAFSNNRKKX
130 140 150 160

```

Based on this analysis, including the gonococcal motifs and the homology with the known PilE protein, it was predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF89-1 (13.6kDa) was cloned in the pGex vector and expressed in *E. coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 11A shows the results of affinity purification of the GST-fusion protein. Purified GST-fusion protein was used to immunise mice, whose sera gave a positive result in the ELISA test., confirming that ORF89-1 is a surface-exposed protein, and that it is a useful immunogen.

Example 41

The following partial DNA sequence was identified in *N. meningitidis* <SEQ ID 345>:

```

1 ATGAAAAAAT CCTCCCTCAT CAGCGCATTG GGCATCGGTA TTTGAGCAT
51 CGGCATGGCA TTTGCCGCC CTGCCGAGCG GGTAAAGCAA ATCCGTCAA
101 ACGCCACTCA AGTATTGAGC ATCTTAAAAA ACGGCGATGC CAACACCGCT
151 CGCAGAAAAA CGAAGCGCTA TGGCATTCCT TATTTCGATT TCCAACGCTAT
201 GACCGCATTG GCGGTCCGCA ACCCTTGGGG CACCG.GTCC GACG.GCAAA
251 AACAGCGGTT GGCCn.AGAA TTCAACCC...

```

This corresponds to the amino acid sequence <SEQ ID 346; ORF91>:

```

1 MKKSLISAL GIGILSIGMA FAAPADAVSQ IRQNATQVLS ILKNGDANTA
51 RQKAEAYAI P YDFQRM TAL AVGNPWXTS DXQKQALAKE FQP...

```

Further work revealed the complete nucleotide sequence <SEQ ID 347>:

```

1 ATGAAAAAAT CCTCCCTCAT CAGCGCATTG GGCATCGGTA TTTGAGCAT
51 CGGCATGGCA TTTGCCGCC CTGCCGAGCG GGTAAAGCAA ATCCGTCAA
101 ACGCCACTCA AGTATTGAGC ATCTTAAAAA ACGGCGATGC CAACACCGCT
151 CGCAGAAAAA CGAAGCGCTA TGGCATTCCT TATTTCGATT TCCAACGCTAT
201 GACCGCATTG GCGGTCCGCA ACCCTTGGGG CACCGCGTCC GACGCGCAAA
251 AACAGCGGTT GGCCAAAGAA TTCAACCC TGCTGATCCG CACCTATTC
301 GGCAGATGCG TGAAATTAAG AAGCGCAAC GTCAACGTCA AAGACAATCC
351 CATCGTCAAT AAGAGCGGCA AAGAAATCAT CGTCCGCGCC GAAGTCGGCG
401 TACCCGGGCA AAAACCGGTC AACATGGACT TCACCACTCA CCAAGCGGCG
451 GTTAAATACC GTACCTACAA CGTGGCCATC GAAGCGCGCA GCGTGGTAC
501 CGTGTACCGC AACCAATTGG GCGAATATAT CAAGCGGAAA GCGGTGGACG
551 GACTGATTGC CGAGTTGAAA GCCAAAAACG GCGGCAATA A

```

This corresponds to the amino acid sequence <SEQ ID 348; ORF91-1>:

```

1 MKKSLISAL GIGILSIGMA FAAPADAVSQ IRQNATQVLS ILKNGDANTA
51 RQKAEAYAI P YDFQRM TAL AVGNPWXTS DXQKQALAKE FQTLLIRTYS
101 GTMLKLKNN VNVKDNPIVN KGGKELIIRA EVGVFGQKPV NMDTITYS
151 GKYRTYNVAI EGASLTVYR NQFGEIKAK GVDGLIAELK AKNGGX*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N. meningitidis* (strain A)

ORF91 shows 92.4% identity over a 92aa overlap with an ORF (ORF91a) from strain A of *N. meningitidis*:

		10	20	30	40	50	60
40	orf91.pep	MKKSLISALGIGILSIGMAFAAPADAVSQIRQNATQVLSILKNGDANTARQKAEAYAI					
	orf91a	MKKSSPISALGIGILSIGMAFAAPADAVNQIRQNATQVLSILKSGDANTARQKAEAYAI					
		10	20	30	40	50	60
45	orf91.pep	YDFQRM TALAVGNPWXTSDXQKQALAKEFQP					
	orf91a	YDFQRM TALAVGNPWXTASDAQKQALAKEFQTLLIRTYS GTMLKLKNNVNVKDNPIVN					
		70	80	90	100	110	120

orf91a KGGKEIIVRAEVGPQGKPVNMDFTTYQSGGKYRTYNVAIEGASLVTYVRNQFGEIIKAK
130 140 150 160 170 180

The complete length ORF91a nucleotide sequence <SEQ ID 349> is:

```

5      1 ATGAAAAAAT CCTCCTTCAT CAGCGCATTG GGCATCGGTA TTTTGAGCAT
51     CGGCATGGCA TTGCGCGCCC CTGCGGACGC GGTAAACCAA ATCCGTCRAA
101    ACGCCACTCA AGTATTGAGC ATCTTAAAAA GCGGTGATGC CAACACCGCC
151    CGCCAAAAGG CCGAAGCCTA TGGGATTCCT TATTTCGATT TCCACGCTAT
201    CGCCGCAATT GCGGTGCGCA ACCCTTGGCC GCGCGGTGCC GACGCGCAAA
251    AACACGCTT GCGCAAGAA TTTCAACCC TGCCTGATCC CACCTATTCG
301    GCGCGATGCG TGAATTTAAA AAACGCGAAC CTCACGCTCA AAGACAATCC
351    CATCGTCAAT AAAGCGGGCA AAGAAATCAT CGTCCGCGCC GAATCGCGCG
401    TACCCGCGGCA AAAACCCGTC AACATGGACT TCACCACTTA CCAAGCGCGC
451    GGTAAATACC GTACCTACAA CGTCCGCAAT GAAGGCGCGA GCCTGGTTAC
501    CGTGTACCGC AACCAATTGC GCGAAATTAT CAAAGCGAAA GGCCTGGACG
551    GACTGATTGC CGAGTTGAAG GCTAAAAACG GCAGCAAGTA A

```

This encodes a protein having amino acid sequence <SEQ ID 350>:

```

1      MKKSSFISAL GIGILSIGMA FAAPADAVNQ IRQNTQVLS ILKSGDANTA
51     RQKAEAYAI P YDFQRM TAL AVGNPWRTAS DAQKQALAKE FOTLLIRIYS
101    GTMLKLNAN VNVKDNPIVN KGGKEIIVRA EVGVPQGKPV NMDFTTYQSG
151    GKYRTYNVAI EGASLVTYVR NQFGEIIKAK GVDGLIAELK ARNGSK*

```

ORF91a and ORF91-1 show 98.0% identity in 196 aa overlap:

```

25      orf91a.pep MKKSSFISALGIGILSIGMAFAAPADAVNQIRQNTQVLSILKSGDANTARQKAEAYAI
           10      20      30      40      50      60
           orf91-1 MKKSSLISALGIGILSIGMAFAAPADAVSQIRQNTQVLSILKNGDANTARQKAEAYAI
           10      20      30      40      50      60

30      orf91a.pep YDFQRM TALAVGNPWRTASDAQKQALAKEFOTLLIRIYS GTMLKLNANVNVKDNPIVN
           70      80      90      100     110     120
           orf91-1 YDFQRM TALAVGNPWRTASDAQKQALAKEFOTLLIRIYS GTMLKLNANVNVKDNPIVN
           70      80      90      100     110     120

35      orf91a.pep KGGKEIIVRAEVGPQGKPVNMDFTTYQSGGKYRTYNVAIEGASLVTYVRNQFGEIIKAK
           130     140     150     160     170     180
           orf91-1 KGGKEIIVRAEVGPQGKPVNMDFTTYQSGGKYRTYNVAIEGASLVTYVRNQFGEIIKAK
           130     140     150     160     170     180

40      orf91a.pep GVDGLIAELKARNGSKX
           190
           orf91-1 GVDGLIAELKARNGSKX
           190

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF91 shows 84.8% identity over a 92aa overlap with a predicted ORF (ORF91.ng) from *N. gonorrhoeae*:

```

50      orf91.pep MKKSSLISALGIGILSIGMAFAAPADAVSQIRQNTQVLSILKNGDANTARQKAEAYAI 60
           orf91.ng VKKSSFISALGIGILSIGMAFASPADAVGQIRQNTQVLTILKSGDAASARPKAEAYAVP 60

55      orf91.pep YDFQRM TALAVGNPWXTSDXQKQALAKEFPQ 93
           orf91.ng YDFQRM TALAVGNPWRTASDAQKQALAKEFOTLLIRIYS GTMLKFNATVNVKDNPIVN 120

```

The complete length ORF91ng nucleotide sequence <SEQ ID 351> is predicted to encode a protein having amino acid sequence <SEQ ID 352>:

```

1 VKKSSFISAL GIGILSIGMA FASPADAVGO IRQNATQVLT ILKSGDAASA
51 RFKAEAYAVP YDFQRMNTAL AVGNFWRTAS DAQKQALAKE FQTLIRITYS
101 GTMLKFKNAT VNVKNDPIVN KGGKEIVVRA EVGIPGQKPV NMDFTTYQSG
151 GKYRTYNVAI EGTSLVTVYR NQFGEIIRAK GIDGLIAELK AKNGGK*

```

5 Further work revealed the complete nucleotide sequence <SEQ ID 353>:

```

1 ATGAAAAAAT CCTCCTTCAT CAGCGCATTG GGCATCGGTA TTTTGAGCAT
51 CGGCATGGCA TTTGCTCTCC CGCGCGACGC AGTGGGACAA ATCCGCCAAA
101 AGGCCACACA GGTFTTGACC ATCTCCAAAA GGGGGGACGC GGCTCTGTGA
151 CGCCCAAAAG CCGAGGCTTA TCCGCTGGCA TATTTCGATT TCCACAGTAT
201 GAGCGATTG CCGGATTCG ACCCTTGGCG TACCGGCTGC GACCGCAAA
251 AACAGCGTIT GGCACAGAA TTTCAAAACC TGCTGATCCG CACCTATTCC
301 GGCACGATGC TGAATTCAA AAACGGGACC GTCAACGTCA AAGACAATCC
351 CATCGTCAAT AAGGGGCGCA AGGAAATCGT CGTCCGTGCC GAAGTCGGCA
401 TCCCGGTTCA GAAGCGCGTC AATATGGACT TTACCACTCA CCAAAGCGGC
451 GGCAATATCC GTACCTACAA CGTGCCTATC GAAGGCACGA GCTCGTTCAC
501 CGTGTACCGC AACCAATTGC GCGAAATCAT CAAAGCCAAA GGCATCGACG
551 GGCTGATTGC CGAGTTGAAA GCCAAAACG GGGCAAAATA A

```

This corresponds to the amino acid sequence <SEQ ID 354; ORF91ng-1>:

```

1 MKKSSFISAL GIGILSIGMA FASPADAVGO IRQNATQVLT ILKSGDAASA
51 RFKAEAYAVP YDFQRMNTAL AVGNFWRTAS DAQKQALAKE FQTLIRITYS
101 GTMLKFKNAT VNVKNDPIVN KGGKEIVVRA EVGIPGQKPV NMDFTTYQSG
151 GKYRTYNVAI EGTSLVTVYR NQFGEIIRAK GIDGLIAELK AKNGGK*

```

ORF91ng-1 and ORF91-1 show 92.3% identity in 196 aa overlap:

```

25 orf91-1.pep      10      20      30      40      50      60
orf91ng-1         10      20      30      40      50      60
MKKSSSLISALGIGILSIGMAFASPADAVSGIRQNATQVLSILKNGDANTARQKAEAYAVP
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
MKKSSFISALGIGILSIGMAFASPADAVSGIRQNATQVLTILKSGDAASARPKAEAYAVP
10      20      30      40      50      60

30 orf91-1.pep      70      80      90      100     110     120
orf91ng-1          70      80      90      100     110     120
YDFQRMNTALAVGNFWRTASDAQKQALAKEFQTLIRITYSGTMLKLNKNDPIVN
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
YDFQRMNTALAVGNFWRTASDAQKQALAKEFQTLIRITYSGTMLKFKNATVNVKNDPIVN
70      80      90      100     110     120

35 orf91-1.pep      130     140     150     160     170     180
orf91ng-1          130     140     150     160     170     180
KGGKEIVVRAEVGIPGQKPVNMDFTTYQSGGKYRTYNVAIEGASLVTVYRNQFGEIIRAK
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
KGGKEIVVRAEVGIPGQKPVNMDFTTYQSGGKYRTYNVAIEGTSLVTVYRNQFGEIIRAK
40      130     140     150     160     170     180

orf91-1.pep      190
GVDGLIAELKAKNGGKX
|:|||||:|||||:
45 orf91ng-1      190
GIDGLIAELKAKNGGKX
190

```

In addition, ORF91ng-1 shows homology to a hypothetical *E. coli* protein:

```

sp|P45390|YRBC_ECOLI_HYPOTHETICAL 24.0 kD PROTEIN IN MURA-RPON INTERGENIC
REGION PRECURSOR (F211).>gi|606130 (U18997) ORF_f211 [Escherichia coli]
>gi|1789583 (AE000399) hypothetical 24.0 kD protein in murZ-rpoN intergenic
region [Escherichia coli]|length = 211
Score = 70.6 bits (170), Expect = 6e-12
Identities = 42/137 (30%), Positives = 76/137 (54%), Gaps = 6/137 (4%)
Query: 59 VFYDFQRMNTALAVGNFWRTASDAQKQALAKEFQTLIRITYSGTMLKFKNATVNVKNDPI 118
+PY + AL +G +++A+ AQ++A F+ L + Y + + T + P
Sbjct: 65 LFYVQKYAGALVLGGYYKSATPAQREAYFAAFREYLRQAYGQALAMHYGQTYQIA--PE 122

Query: 119 VNMKGKEIV-VRAEVGIP-GQKPVNMDFTTYQSG--GKYRTYNVAIEGTSLVTVYRNQF 174
G K I V + B P G + P Y + D F + + G + + Y + + E G S + T + N + G
Sbjct: 123 QPLGDRITIVPIRVTIIDFNRPGRVRLDFQWRKNSQTGNQWQAYDMIAEGVSMITTKQNEW 182

```


-232-

		70	80	90	100	110	120
orf97.pep		MDIFAVIDHCEAARRNGLTMQPAKVI	VFGT	PKAGT	PLMWKDP	PAFALQLPLRV	LVETD
orf97a		MDIFAVIDHCEAARRNGLTMQPAKVI	VFGT	PKAGT	PLMWKDP	PAFALQLPLRV	LVETD
		70	80	90	100	110	120
		130	140	150	160		
orf97.pep		VRAAYTDTRALIAGSRIGFDEVANT	LANAEKLI	QKT	VGEX		
orf97a		VRAAYTDTRALIAGSRIGFDEVANT	LANAEKLI	QKT	IGEX		
		130	140	150	160		

The complete length ORF97a nucleotide sequence <SEQ ID 359> is:

	1	ATGANACACA	TACTCCOCC	T	GANTGNC	GCA	TCOGCACT	GCATTTC	CAAC
	51	CGCTTCGGNN	CATCCTGCCA	GCGAACG	CGCA	AAACCAAA	AC	GAACCG	CGCTA
	101	TGACCACGCA	TACCTCACC	TCAAAAT	ACA	GTTTTCAG	CA	AACCGT	CAGC
	151	CGOCTTGAAA	COGCCATA	AAAGCAAA	AGGG	ATGGACAT	TTT	TGCGCT	CAT
	201	CGACCATCAG	GAGCCGCG	CCGCCAA	ACGG	CTTAACG	ATG	CAGCCG	CGCAA
	251	AAGTCACTGT	CTTGGCG	ACG	CCCAAG	CGCG	GTAACG	CGCT	GATGGT
	301	GACCCCGCT	TGCGOCT	GCA	ACTGCC	CCCTG	CGGTC	NTG	CTACG
	351	GGACGGCAAA	GACGCGCG	CG	CTATAC	CGGA	TACGCG	CGCG	CTCAT
	401	GCAGCGCAT	CGGTTTC	GAC	GAGTGG	CA	ACATTT	TGGC	AAACG
	451	AAACTGATC	AAAAACCA	T	AGGCGA	ATAA			

This encodes a protein having amino acid sequence <SEQ ID 360>:

	1	MXHILPLXXA	SALCISTASX	HPASEPOTN	ETAMITH	TLT	SKYSFDET	VS
	51	RLETAIKSKG	MDIFAVIDH	CEAARRNGL	TMQPAKVI	VFGT	PKAGT	PLMWK
	101	DPAFALQLPL	RVXVETD	GK	VRAAYTD	TRALIAGSRIG	FDEVANT	LANAE
	151	KLIQKTIGE*						

ORF97a and ORF97-1 show 95.6% identity in 159 aa overlap:

		10	20	30	40	50	60
orf97a.pep		MXHILPLXXASALCISTASXHPASEPOTN	ETAMITH	TLT	SKYSFDET	VS	RLETAIKSKG
orf97-1		MXHILPLIAASALCISTASAHFASEPSTQNETAM	TH	TLT	SKYSFDET	VS	RLETAIKSKG
		10	20	30	40	50	60
		70	80	90	100	110	120
orf97a.pep		MDIFAVIDHCEAARRNGLTMQPAKVI	VFGT	PKAGT	PLMWKDP	PAFALQLPLRV	LVETD
orf97-1		MDIFAVIDHCEAARRNGLTMQPAKVI	VFGT	PKAGT	PLMWKDP	PAFALQLPLRV	LVETD
		70	80	90	100	110	120
		130	140	150	160		
orf97a.pep		VRAAYTDTRALIAGSRIGFDEVANT	LANAEKLI	QKT	IGEX		
orf97-1		VRAAYTDTRALIAGSRIGFDEVANT	LANAEKLI	QKT	VGEX		
		130	140	150	160		

Homology with a predicted ORF from *N. gonorrhoeae*

ORF97 shows 88.1% identity over a 159aa overlap with a predicted ORF (ORF97.ng) from *N. gonorrhoeae*:

50	orf97.pep	MXHILPLIAASALCISTASAHFASEPSTQNETAMITH	TLT	SKYSF	DET	VS	RLETAIKSKG	60
	orf97ng	MXHILPLIAASALCISTASAHFASEPSTQNETAMITH	TLT	SKYSF	DET	VS	RLETAIKSKG	60
55	orf97.pep	MDIFAVIDHCEAARRNGLTMQPAKVI	VFGT	PKAGT	PLMWKDP	PAFALQLPLRV	LVETD	120
	orf97ng	MDIFAVIDHCEAARRNGLTMQPAKVI	VFGT	PKAGT	PLMWKDP	PAFALQLPLRV	LVETD	120
	orf97.pep	VRAAYTDTRALIAGSRIGFDEVANT	LANAEKLI	QKT	IGEX			159
	orf97ng	VRAAYTDTRALIAGSRIGFDEVANT	LANAEKLI	QKT	IGEX			159

The complete length ORF97ng nucleotide sequence <SEQ ID 361> is predicted to encode a protein having amino acid sequence <SEQ ID 362>:

```

1  MKHILPPIAA  SAFCISTASA  HPAGKPPTON  ETAMTHTLTL  SKYSFDETVS
51  RLETAIKSKG  MDIFAVTDHG  EAARRNGLTM  QPAKIVFEGT  PKAGTPLMVK
101 DPAPALQLPL  RVLVTETDGG  VRTAYTDTRA  LIVGSRISFD  EVANTLANAE
151 KLIQKTVGE*
```

Further work revealed the complete nucleotide sequence <SEQ ID 363>:

```

1  ATGAAACACA  TACTCCCGCT  gatcgccgca  TccgactCT  GCATTTCAAC
51  CGGCTTCGGCA  CACCCTCGCG  GCAAAACGCG  CACCCAAAAC  GAAACCGGTA
101 TGACCCAGCGA  CACCCTCACC  TCGAAATACA  GTTTTGACGA  AACCGTCAGC
151  CGCCTTGAAA  CGGCCATAAA  AAGCAAAGGG  ATGGACATT  TTGCCGTCAT
201  CGACCATCAG  GAAGCGGCAC  GCGCAAAGCG  CCTGACCATG  CAGCGCGCAA
251  AAGTCATCGT  CTTGCGCACG  CCCAAGGCGG  GTACGCGCGT  GATGCTCAAA
301  GACCCCGCCT  TCGCCCTGCA  ACTGCGCCGT  CGCGTCCTGT  TTACCGAAAC
351  GGACGGCAAA  GTACGCGACG  CTTATACCGA  TACCGCGCGC  CTCATCGTGT
401  GCAGCCGCAT  CAGTTTCGAC  GAAGTGGCAA  ACACCTTGGC  AAACGCCGAA
451  AAAGTGATAC  AAAAAACCGT  AGGCGATRA
```

This corresponds to the amino acid sequence <SEQ ID 364; ORF97ng-1>:

```

1  MKHILPLIAA  SALCISTASA  HPAGKPPTON  ETAMTHTLTL  SKYSFDETVS
51  RLETAIKSKG  MDIFAVTDHG  EAARRNGLTM  QPAKIVFEGT  PKAGTPLMVK
101 DPAPALQLPL  RVLVTETDGG  VRTAYTDTRA  LIVGSRISFD  EVANTLANAE
151 KLIQKTVGE*
```

ORF97ng-1 and ORF97-1 show 96.2% identity in 159 aa overlap:

		10	20	30	40	50	60
25	orf97-1.pep	MKHILPLIAASALCISTASAHFASEFSTQNETAMTHTLTSKYSFDETVSRLETAIKSKG					
	orf97ng-1	MKHILPLIAASALCISTASAHFAGKPKPTQNETAMTHTLTSKYSFDETVSRLETAIKSKG					
		10	20	30	40	50	60
30	orf97-1.pep	MDIFAVTDHQAARRNGLTMQPAKIVFEGTPKAGTPLMVKDPAPALQLPLRVLVTETDGG					
	orf97ng-1	MDIFAVTDHQAARRNGLTMQPAKIVFEGTPKAGTPLMVKDPAPALQLPLRVLVTETDGG					
		70	80	90	100	110	120
35	orf97-1.pep	VRAAYTDTRALIAGSRIGFDEVANTLANAEKLIQKTVGEX					
	orf97ng-1	VRTAYTDTRALIVGSRISFDEVANTLANAEKLIQKTVGEX					
		130	140	150	160		
40	orf97-1.pep	VRAAYTDTRALIAGSRIGFDEVANTLANAEKLIQKTVGEX					
	orf97ng-1	VRTAYTDTRALIVGSRISFDEVANTLANAEKLIQKTVGEX					
		130	140	150	160		

Based on this analysis, including the presence of a putative leader sequence in the gonococcal protein, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF97-1 (15.3kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figures 12A & 12B show, respectively, the results of affinity purification of the GST-fusion and His-fusion proteins. Purified GST-fusion protein was used to immunise mice, whose sera were used for Western Blot (Figure 12C), ELISA (positive result), and FACS analysis (Figure 12D). These experiments confirm that ORF97-1 is a surface-exposed protein, and that it is a useful immunogen.

Figure 12E shows plots of hydrophilicity, antigenic index, and AMPHI regions for ORF97-1.

Example 43

The following DNA, believed to be complete, sequence was identified in *N.meningitidis* <SEQ ID 365>:

5	1	ATGGCTTTTA	TTACGGCGTT	ATTCAAAAGC	AGTAATAGGC	TSATTGTGGC
	51	CGTATGCTTC	CGCGCGTTTC	AGATGTGTGG	GGCGAGAGGG	ATGATGTGGC
	101	CGCGTGGCGA	ACGAGAGATA	ACCGACGGCG	GGAGCACTTC	CATGACGAGA
	151	CGCTTCCAAA	CCGACGTGGC	CGAACCATTC	CACACCGGTC	TGCGCGAGGG
10	201	CGTGGTGGCG	ATTCTTAACT	AGTCTGGCA	CGTTTCCGCG	CCGATTAATCG
	251	CTTCTTATCG	GTTTAACTAG	GGGCACTAGA	TGGCGGACGA	CGACATATT
	301	CGACTACAAA	TGAGTTTCCA	TCCGCTGACC	AAACGCTACC	CGGTACTCGT
	351	CGGCGCGGTT	TGCACAGACT	ACGACCGATT	GGATGGAGCT	TGGCGCGCGA
	401	CGCGCGCGGT	TGCGCACTGC	AAAGTGTAGA	ACAAAGGCGG	GCTGTCCGGT
	451	CGGGAGACGA	GGGAACACAA	GGCGCGCGAT	CGCGTGAAGC	TGTCCTCTTC
15	501	AAATCTGCCC	AGGCTTTTTC	AAATCACTGC	ATTGACGATC	CAAAATCTGC
	551	ATTTGGATTC	GGGTTGGAAA	CTCTTAARCA	TCATCTGGGA	CAATAAA

This corresponds to the amino acid sequence <SEQ ID 366; ORF106>:

20

```
1  MAFITRLFKS SKWLIVPLML PAFQNVAAEG IDVSRAEARI TDGGQLSISS
51 RFQTELPDQL QOALRGVPL NTFLSWOLSA PIISYRFLK QGLIGDDDDN
101 DYKLSFHPLT KRYRVTVGAF STDYDTLDA LRA TGAVANV KVLNKGALSG
151 AEAGETKAEI LRLTSLKPL KPFQINLDS QNWHLDSGWK PLNIIGNK*
```

Further work revealed the following DNA sequence <SEQ ID 367>:

25	1	ATGGCTTTT	TTCCGCGCTT	AGTAAAGAG	AGTAAATGGC	TGATGTGGCC
	51	GTGATGATG	CACCGCTTTC	AGTAATGGGC	GGCGGAGGGG	ATAGATGTGC
	101	GGCGTCCGGA	ACCGAGGATTA	ACCGAGGAGG	GGCGACTTTC	CATCAGCAGC
	151	CGCTTCCAAA	CGCGATGTGC	CGACCACTTC	CACACGGTTC	GGCGCGGGG
	201	CGCTGCGATC	ACATTTACTT	TACGTAGGCA	CGCTTCCGCG	CGGATATGCG
30	251	CTCTCTATGT	CGGACCTAGA	CGGACCTAGA	CGGACCTAGA	CGGACCTAGA
	301	GACTACAAAC	TGGTTTTCGA	TGGCTGACCC	ACCGCTGACT	GGGTATCGGT
	351	CGGCGCGTPT	TGCGACAGCT	AGCAACCTTT	TGGGCGCGCA	TGGCGCGCGA
	401	CGGCGCGGPT	TGCGCAACTG	AAAGCTCTGA	ACAAAGAGCG	CGTGTGCGPT
	451	CGGAGCAGAC	GGGAAACCAAC	GGGGAATATC	CGCGTACGCG	TGTCCACTTC
35	501	GAACCTCTTC	GAATCAACGC	GAATCAACGC	GAATCAACGC	GAATCAACGC
	551	ATTTCCTATC	GGGTTTAAAC	GGGTTTAAAC	GGGTTTAAAC	GGGTTTAAAC
	601					

35 This corresponds to the amino acid sequence <SEQ ID 368; ORF106-1>:

```

1  MAFITRLFKS SKWLIPLML PAFQNVAAEG IDVSRAEARI TDGGQLSISS
51 RFQTELPDQL QQALRRGVP LNTLSWQLSA PIATSYRFKL GQLIGDDDDN
101 DYKLSFHEPL RRYVTVFGL STDYDTLDAA LRATGADVSN KVLNKGALSG
151 AEGETKAEK NLTYSKSLP KPFQINALTS ONWHLDGSKW PLNTIGNK*

```

40 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF106 shows 87.4% identity over a 199aa overlap with an ORF (ORF106a) from strain A of *N. meningitidis*.

[illegible]

-235-

orf106a	LQXAXXRGVXLNKTLXWQLSAPIIASYRFXLGLIGDDDXIDYKLSFHLPLTNRYRVTVGA	70	80	90	100	110	120
5	orf106.pep	120	130	140	150	160	170
	orf106a	180	190	199			
10	orf106.pep	180	190	199			
	orf106a	180	190	200			
15	Due to the K→N substitution at residue 111, the homology between ORF106a and ORF106-1 is 87.9% over the same 199 aa overlap.						

The complete length ORF106a nucleotide sequence <SEQ ID 369> is:

1	ATGGCTTTTA	TTACGCGCTT	ATTCAAAGC	ATTAAACAAT	GGCTTGCTGCT
51	GCTGCCGATG	CTTTCCGTTT	TGCCGAGCG	GGCGCGGAG	GGGATAGATG
101	TGAGCGCGCG	CGAAGCGAGG	ATAACGACG	CGCGGACAGT	TTCCATNAGN
151	AGCGCGCTCC	AAACCGAGCT	GCCGACGAG	CTCCANNNG	CGNNGCGCG
201	GGCGGTGNCG	CTCAACTNTA	CCTTAAGNTG	GCAGCTTCC	GCCCCGATAA
251	TGCTCTCTTA	TCGGTTTNA	TTGGGGCAAC	TGATTGGCGA	TGACGACNAT
301	ATTGACTACA	AACGTGATTT	CCATCCGCTG	ACCAACCGCT	ACCGCGTTAC
351	CGTCGGCGCG	TTTTCGACAG	ATAGACAC	CTTGATGGG	GCATTGGCG
401	CGACCGCGCG	GGTTGCCAAC	TGGAAGTCC	TGACAAAGG	CGCGCTGCC
451	GGTGGCGAAG	CAGGGGAAAC	CAGGCGGAA	ATCCGCGTGA	CGCTGTCCAC
501	TTCAAAACTG	CCCAAGCCTT	TTCAAACTCA	TGCATTGACT	TCTCAAAACT
551	GGCATTGGGA	TTCCGGTTGG	AAACCTCTAA	ACATCATCGG	GAACAAATAA

This encodes a protein having amino acid sequence <SEQ ID 370>:

1	MAFITRLFKS	IKQVLVLLPM	LSVLPDAAAE	GIDVSRAREAR	IXDGGQLSXX
51	SRFQTELPDQ	LQXAXXRGVX	LNKTLXWQLS	APIIASYRFX	LGQLIGDDDX
101	IDYKLSFHL	TNRYRVTVGA	FSTXYDTLDA	ALRATGAVAN	WKVLNKGALS
151	GAEAGETKAE	IRLTLSKSL	PKFPQINALT	SQNWHLDSGW	KPLNIIGNK*

Homology with a predicted ORF from *N.gonorrhoeae*

ORF106 shows 90.5% identity over a 199aa overlap with a predicted ORF (ORF106.ng) from *N.gonorrhoeae*:

40	orf106.pep	MAFITRLFKSSK-WLIVPLMLPAFQNVAAEGIDVSRAREARITDGGQLSISRRFQTELPDQ	59
	orf106ng	MAFITRLFKSIKQVLVLLPILSVLPDAAAGIAATRAEARITDGGRLSISRRFQTELPDQ	60
	orf106.pep	LQALRRGVPLNFTLSWQLSAPIIASYRFXLGLIGDDDDNIDYKLSFHLPLTKRYRVTVGA	119
45	orf106ng	LQALRRGVPLNFTLSWQLSAPIIASYRFXLGLIGDDDDNIDYKLSFHLPLTNRYRVTVGA	120
	orf106.pep	FSTDYDTLDAALRATGAVANWKVLNKGALSAGAEAGETKAEIRLTLSLSKLPKFPQINALT	179
50	orf106ng	FSTDYDTLDAALRATGAVANWKVLNKGALSAGAEAGETKAEIRLTLSLSKLPKFPQINALT	180
	orf106.pep	SQNWHLDSGWKPLNIIGNK	198
	orf106ng	SQNWHLDSGWKPLNIIGNK	199

Due to the K→N substitution at residue 111, the homology between ORF106ng and ORF106-1 is

91.0% over the same 199 aa overlap.

The complete length ORF106ng nucleotide sequence <SEQ ID 371> is:

```

1  ATGGCTTTTA TTACGCGCTT ATTCAAAGC ATTAACAAT GGCTTGTGCT
51  CTCTCCGATA CTCTCCGCTT TGCCGAGCG GCGCGGAGG GGCATTGCGG
101  GACCCGCGCG CGAAGCGAGG ATAACCGAG CGGCGCGCTT TTCCATCGCG
151  ACCGCGCTCC AAACCGAGCT GCCCGACGAC CTCGACAGCG GCTTGGCGCG
201  GGGCGTACCG CTCACCTTTA CTTTAAAGCT GCGAGCTTTC GCCCGACAAA
251  TTGCTTCTTA TCGGTTTAAA TTGGGGCAAC TGATTGGCGA TGACGACAA
301  ATTGACTACA AACTAAGTTT CCATCCGCTG ACCAACCGCT ACOCGCTTAC
351  CCGCGCGGCA TTTTCCACCG ATTACGACAC TTTGGATGCG GCATTGCGCG
401  CGACCGGCGC GGTTCGCAAC TGGAAAGTCC TGAACAAAGG CGCGTTGTCC
451  GGTGCGGAAG CAGGGGAAAC CAGGGGCGAA ATCCGCGCTGA CGCTGTCCAC
501  TTCAAAACCT GCCAAGCCTT TCCAAATCAA CGCATTGACT TCTCAAAACT
551  GGCATTGGA TTCGGGTTGG AAACCTCTAA ACATCATGCG GAACAAATAA

```

This encodes a protein having amino acid sequence <SEQ ID 372>:

```

1  NAFITRLFKS IKQWLVLLPI LSVLPDAAAE GIAATRAEAR ITDGGRLSIS
51  SRFOTELPDQ LQALRRGVP LNFITLWQLS APTIASYREK LQQLIGDDNN
101  IDYKLSFHL TNRVRYTVGA FSTDYDYLDA ALRATGAVAN WKVINKGALS
151  GARAGETHAE IRLTLSTSKL PKFQINALT SQNWLDSGW KPLNITGNK*

```

Based on this analysis, including the presence of a putative leader sequence in the gonococcal protein, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF106-1 (18kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 13A shows the results of affinity purification of the His-fusion protein, and Figure 13B shows the results of expression of the GST-fusion in *E.coli*. Purified His-fusion protein was used to immunise mice, whose sera were used for FACS analysis (Figure 13C) These experiments confirm that ORF106-1 is a surface-exposed protein, and that it is a useful immunogen.

Example 44

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID 373>:

```

1  ATGGACACAA AAGAAATCCT CGG.TACGCG GcAGGcTGA TOGGCAGCGC
51  GGTFTTAGCG GTCAATCATCC TGCCGCTGCT GTGCTGGTAT TTCCCCGCGG
101  ACGACATCGG GCGCATCGTG CTGATGACGA CGCGCGCGGG GCTGACGGTG
151  TCGGTGTGTT GCCTCGGGCT GGATCAGGCA TACGTCGCGG AATACTATGC
201  CACCGCGGAC AAAGACAaCT TGTCAAAAC CCGTGTCTCG CGCGCGCTGC
251  TGTCTGCGCG CGGATAGGCC GCCCTGCTGC TTTCCGCGCC GTCCCTGCGG
301  TCTGAATATCC TGTTTTCRCT CGACGATGCC GCGCGCGGCa TCGGGCTGGT
351  GCTGTTTGAa CcGAGCTTCC TGCCCATCCG CTTCCTCTTA CTGGTFTTGC
401  GTATGGAAGG ACGGCGcCTT GCGTFTTCTG CGCGGCACT CTGCGCGAG
451  CTCGCGATCG TGCGTCTG.T GCGGCTGAGC GTGCGGCTCG TGCACTTCC
501  AGCGAACACC GCGCTCCTGA CCGCGCTTAA CGCGCTGGCA AACCTTGGCG
551  CCGCGCGCTT TTTGCTGTTT CAAACCGGAT GCGCTCTGAA GGCGCTCCGG
601  CACGCAACCT TTTGCGCCCG CGTCTGCGAC GCGGGG.TGC GCTACGGCAT
651  ACCGATCGCA CTGAGCAGCA TCGCCTATTG GGGGCTGGCA TCCGCGGACC
701  GTTTGTTCCT GAAAAATAT GCGCGCCTGG AACAGCTCGG CTTTATTTCG
751  ATGGGATATT CGTTCGGCGG GCGGCAATTA TTGTTCCAAA GATCCTTTTC
801  AACGCTCTGG ACACGTTATA TTTTCCGCGC AATCGAAGAA AACGCGCCGC
851  CCGCTCGCCT CTCGGCAACG GAGAAATCGG CGCGCGCCCT GCTTGCTCC
901  GCGCTCTGc. TGACGCGCAT TTTCTGCGCC CTTCGCTGCC TCTGCTGCC
951  GGAAAACTAC GCGCGCGTCC GGTTTATCGT ATG.TGCGCG

```

5
 1001 CGCTGTTTTG CACGCTGGCG GAAATCAGCG GCATCGSTTT GAACGCTGTT
 1051 CGCAAAACGC GCCCGATCGC GCTCGCCACC TTGGGCGCGC TGGCGGCAAA
 1101 COTGCTGCTG CTGGGGCTTG ACCGTGCCGT ACCGCGAGG CGCGC.GGCG
 1151 CGCGGCTTGC CTGTCGCCGC TCATTCTGGC TGTTTTTTTC CTCTCAAGACC
 1201 GAAAGCTCyT GCCGCGCTGC GCAGCGGCTC AAACGCGCTCG CGCTTTATCT
 1251 GCACACATTG TTCTGCTGTA CCTCTCTGCG GGCGCTACCC TGCTTCGGCA
 1301 CGCGGGCAAA CTATCCCGCT TTTGCGGCGC TATGGCGGCA ATATCTGGCA
 1351 GGTGTCATCC TCGGCCACCG GAAAGATTG CACAAACTGT TTCATTATTT
 1401 GAAAAACAA GTTTCCCAT TATGA

10 This corresponds to the amino acid sequence <SEQ ID 374; ORF10>:

1 MDTKEILXYA AGSIGSAVLA VIILPLLSWY FPADDIGRIV IMQTAAGLTV
 51 SVLCGLDQA YVREYYATAD KDTLFTLFL PPLLSAAIA ALLSRFSLP
 101 SEILFSLDQA AAGIGLVLE LSLPIRFL LVLMEGRAL AFSSAQLVPK
 151 LAIILLPLT VGLHFPANT AVLTAVYALA NLAAAFLLF QNRCRLKVR
 201 HAPFSPAVLH RGXRYGPIA LSSIAWGLA SADRLFLKY AGLEQLGVYS
 251 MGISFGAAL LFQSFSTVW TPYIFRAIEE NAPPARISAT AESAAALLAS
 301 ALCLTGIFSP LASILLPENY AAVRFIVVSC MXPPLFCTLA EISGIGLVV
 351 RKRTRIALAT LGALANLLL LGLDRAVPAF PXGAAVACAA SFWLFFAFKE
 401 ESSCRLWQL KRLPLYLHTL FCLTSSAAYT CFGTANYPLF FAGVWAAYLA
 20 451 GILHRKDLH KHLFYLKQG GFPL*

Further sequence analysis revealed the complete DNA sequence<SEQ ID 375> to be:

1 ATGGACACAA AAGAAATCCT CGGCTACGCG GAGCGCTCGA TCGGACGCGC
 51 GGTTTTGACC GTCATCATCC TCGCGCTGCT GTCTGGTAT TTCCCGCGCG
 101 ACGACATCGG GCSCATCGTG CTGATGCAGA CGCGCGCGCG ACTGACGGTG
 25 151 TCGGTGTTGT GCCTCGGCTG GGATCAGGCA TACGTCCGGG AATACTATGC
 201 CACCGCGCAG AAGACACACT TGTTCAAAC CTGTTTCTCG CGCGCGCTGC
 251 TGCTCGCGCG CGCGATAGCC GCCCTGCTGC TTTCGCGCCC GTCCCTGCGG
 301 TCTGAAATCC TGTTTTCACT CGACGATGCC CGCGCGGCA TCGGGCTGGT
 351 GCTGTTTGAA CTGAGCTTCC TGCCCATCCG CTTTCTCTTA CTGTTTTCG
 30 401 GTATGGAAGG ACGCGCCCTT GCCTTTTCGT CGCGCGAATC CGTGCCCAAG
 451 CTCGCGCATCC TGCTGCTGCT GCGCTGACG GTCGGCGTGC TGCACTTTCC
 501 AGCGAACACC GCGGTCTGA CGCGCGTTA CGCGCTGCA AACCTTGCGG
 551 CGCGCGCTT TTTGCTGTTT CAAAACCGAT GCGGTCTGAA GCGCTCGCG
 601 CACGACCGT TTTGCGCGC GTCCTGCAC CGGCGCTGCG GCTACGGCAT
 35 651 ACGATCGCA CTGAGGACCA TCGCTATTG GCGCTGCA TCCCGGAC
 701 GTTTGTCCT GAAAAATAT CGCGCGCTG AACAGCTCG CGTTTATTCG
 751 ATGGGTATT GTTTCGCGCG GCGCGCATA TTGTTCCAAA GCATCTTTC
 801 AACGCTCTGG ACACCGTATA TTTCCGCGC AATCGAAGAA AACGCGCGC
 40 851 CGCGCGGCT CTCGCGACG GAGAATCGC CGCGCGGCTT TCTGCTGCC
 901 GCGCTCTGCC TGACCGGCAT TTTCTGCGC CTGCGCTCCC TCTGCTGCC
 951 GGAACACTAC GCGCGCGTCC GGTTTATCGT CGTATCGTGT ATGCTGCGCG
 1001 CGCTGTTTTG CACGCTGGCG GAAATCAGCG GCATCGSTTT GAACGCTGC
 1051 CGCAAAACGC GCCCGATCGC GCTCGCCACC TTGGGCGGCG TGGCGGCAAA
 1101 COTGCTGCTG CTGGGGCTTG CGGTGCGCTC CGCGCGCGCG CGCGCGCGG
 45 1151 CGSTTGCGTG TGCGCGCTCA TTCTGCGCTG TTTTTCGCTT CAAGACCGAA
 1201 ACGTCTGCGC GCGCTGCGCA CGCGCTCAA CGCTGCGCG TTATCTGCA
 1251 CACATCTTTC TGCTGACCT CCGCGCGCG CTACACCTCG TCGGCAAGC
 1301 CGCAAACTAT TCGCTGTTT GCGCGCAT TCGCGCGCAT TCGGCAAGC
 1351 TGCTCTGCTG GCCACCGGAA AGATTTCAC AACTGTTCCT ATTTATTGAA
 50 1401 AAAACAAGT TTCCCATAT GA

This corresponds to the amino acid sequence <SEQ ID 376; ORF10-1>:

1 MDTKEILGYA AGSIGSAVLA VIILPLLSWY FPADDIGRIV IMQTAAGLTV
 51 SVLCGLDQA YVREYYATAD KDTLFTLFL PPLLSAAIA ALLSRFSLP
 101 SEILFSLDQA AAGIGLVLE LSLPIRFL LVLMEGRAL AFSSAQLVPK
 55 151 LAIILLPLT VGLHFPANT AVLTAVYALA NLAAAFLLF QNRCRLKVR
 201 HAPFSPAVLH RGLRYGPIA LSSIAWGLA SADRLFLKY AGLEQLGVYS
 251 MGISFGAAL LFQSFSTVW TPYIFRAIEE NAPPARISAT AESAAALLAS
 301 ALCLTGIFSP LASILLPENY AAVRFIVVSC MXPPLFCTLA EISGIGLVV
 351 RKRTRIALAT LGALANLLL LGLAVPSSGA RGAAVACAA SFWLFFAFKE
 60 401 SSCRLWQLK RLPLYLHTL FCLTSSAAYT FGTPANYPLF AGVWAAYLAG
 451 CILHRKDLH KHLFYLKQG FPL*

Computer analysis of this amino acid sequence gave the following results:

Prediction

ORF10-1 is predicted to be the precursor of an integral membrane protein, since it comprises several (12-13) potential transmembrane segments, and a probable cleavable signal peptide

Homology with EpsM from *Streptococcus thermophilus* (accession number U40830).

- 5 ORF10 shows homology with the *epsM* gene of *S. thermophilus*, which encodes a protein of a size similar to ORF10 and is involved in exopolysaccharide synthesis. Other homologies are with prokaryotic membrane proteins:

Identities = (25%)

10 Query: 213 LRYGIPLALSSLAYWGLASADRLFLKKYAGLEQLGVYSMGISFGGAALLQSIFSTVW 270
L Y + PL SS+ + W L + + R F+ + G G+ ++ + + IF+ W
Spict: 210 LYALPLIPSSILWLLNASSRYFVLEFLGAGANGLLAVATKIPSIISIENTIFTORW 267

Identities = 15/57 (26%), Positives = 31/57 (54%)

Query: 7 LGYAAGSIGSAVLAVIILPLLWYFPADDIGRIVLMQTAAGLTVSVLCLGLDQAYVR 63
L + G+GS +L +++PL ++ + G LQTAL + + + + A +R
Sbjct: 12 LVFTIGNLGSKLLVFLVLPYTYAMTPOEYGMADLYOTTANLLLPLITMNVFDTLR 68

Identities = 16/96 (16%), Positives = 36/96 (37%)

Query: 307 IFSPLASLLLPENYA AAVRFTVVS CMLPPLFYTLTEISGIGLNVRKTRPIXXXXXXX 366
+ P+ ++ +YA+ V ML LF + ++ G ++T+ +
25 Spict: 305 VLKPIVEKVVSSDYASSWOYVPEFMLSMLFSSFSDFEGTNYIAAKOTKGVEMTSIYGTIV 364

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF10 shows 95.4% identity over a 475aa overlap with an ORF (ORF10a) from strain A of *N. meningitidis*:

30	orf10.pep	MDTKTEILKYAGSGISGSAVLAVIILLPLLSWFFPADDIGRIVLMQTAAGLTVSVSLCGLDQA	10	20	30	40	50	60
	orf10a	MDTKTEILKYAGSGISGSAVLAVIILLPLLSWFFPADDIGRIVLMQTAAGLTVSVSLCGLDQA	10	20	30	40	50	60
35	orf10.pep	YVREYYATADKDTLTKFLFLPPLLSAAAIQAALLSRPSLPSEILFSLDDAARAGIGLVLFE	70	80	90	100	110	120
	orf10a	YVREYYAAADKDTLTKFLFLPPLLSAAAIQAALLSRPSLPSEILFSLDDAARAGIGLVLFE	70	80	90	100	110	120
40	orf10.pep	LSFLPIRIFLLVLRMEGRALAFSSAQVLKAILLLXPLTVGLLHFPANTAVLTAVYALA	130	140	150	160	170	180
	orf10a	LSFLPIRIFLLVLRMEGRALAFSSAQVLSKAILLLPLTVGLLHFPANTAVLTAVYALA	130	140	150	160	170	180
45	orf10.pep	NLAAAAFLFPQNRCLKXAVRHAPFSSAVLHGRGXRYGPIALSSIAYWGLSADRFLFKKY	190	200	210	220	230	240
	orf10a	NLAAAAFLFPQNRCLKXAVRHAPFSSAVLHGRGXRYGPIALSSIAYWGLSADRFLFKKY	190	200	210	220	230	240
50	orf10.pep	AGLEQLGVYSMGISFGGAALLFQSFISTVWTVPYIFRAIEANAPPARLSATSAESAALLAS	250	260	270	280	290	300
	orf10a	AGLEQLGVYSMGISFGGAALLFQSFISTVWTVPYIFRAIEANAPPARLSATSAESAALLAS	250	260	270	280	290	300

-239-

		310	320	330	340	350	360
orf10.pep		ALCXTGIFSP	PLASLLLPENYA	AVRFIVVSCMXP	FLFCTLA	EISGIGLN	VVRKTRPIALAT
5	orf10a	ALCLTGI	SPPLASLLLPENYA	AVRFIVVSCMLP	FLFCTLVEISGIGLN	VVRKTRPIALAT	
		310	320	330	340	350	360
		370	380	390	400	410	419
orf10.pep		LGALAANLL	LGLDRAV	FAR-FXGA	AAVCAASEWLF	FAFKTESS	CRLWQPLKRLPLYLHT
10	orf10a	LGALAANLL	LGL--AVP	SGGARGA	AAVCAASEWLF	VPFKTESS	CRLWQPLKRLPLYMHT
		370	380	390	400	410	
		420	430	440	450	460	470
orf10.pep		LFCLTSS	AAVTCFGT	PANYP	LFAGVWA	YLAGCILR	HRKDLHKL
15	orf10a	LFCLTSS	AAVTCFGT	PANYP	LFAGVWA	YLAGCILR	HRKDLHKL
		420	430	440	450	460	470

The complete length ORF10a nucleotide sequence <SEQ ID 377> is:

20	1	ATGGACACAA	AAGAAATCCT	CGGCTACGG	CGAGGCTCGA	TCGGCAGCGC
	51	GGTTTATGCC	GTCAATCATCC	TGCGCTGCTG	GTCTGTGTAT	TTCCCTGCGG
	101	ACGACATCGG	AGCATCGCTG	CTGATGCGAG	CGGCGGGCGG	GCTGACGGTG
	151	TCGGTGTGTG	GCCTCGGGCT	GGATCAGCGA	TACGTCGCGG	AGTACTATGC
	201	CGCGCGCGAC	AAGACACATT	TGTTCAAAAC	CTGTTCCTGG	CGCGCGCTGC
	251	TGCTGTCGCG	CGCGATAGCC	GCCTCTGCTG	TTTCCGCGCC	ATCCCTGCGG
25	301	TCGTGAATCC	TGTTTTCGCT	CGACGATGCC	CGCGCGCGCA	TCGGGCTGGT
	351	GCTGTTTGAA	CTGAGCTTCC	TGCCATCCG	CTTTCTCTTA	CTGGTTTTCG
	401	GTATGGAAGG	ACGCGCCCTT	GCCTTTTCGT	CGCGCAACT	CGTGTCCAAG
	451	CTCGCATACC	TGCTGCTGCT	CGCGCTGACG	CTCGGGCTGC	TGCACTTTCC
	501	GGGGAACACC	GCGCTCCTGA	CGCGCTTTTA	CGCGCTGGCA	AACCTTGCGG
30	551	CGCGCGCCTT	TTTGCTGTTT	CAAAAACGAT	CGCGCTGAGA	GGCGCTCCGG
	601	CGCGCACCGT	TTTCATCCGC	CGTCTGCTAT	CGCGGCTCGC	GCTACGGCAT
	651	ACCGATCGCA	CTAAGCAGCA	TGCGCTATTG	GGGGCTGGCA	TCGCGCGACC
	701	ATTTGTTCTC	GA AAAAATAT	CGCGGCTAG	AACAGCTGGG	GCTTTATTCG
	751	ATGGGTGTTT	CGTTCGCGCG	ACGGGCAATA	TTTGTTCGAA	GCATCTTTTC
35	801	AACGGTCTGG	ACACCGCTATA	TTTTCCGCGC	AATGAGAGCA	AAGCGCCGCG
	851	CGCGCCGCGT	CTCGGCAACG	CGAGATCCG	CGCGCGCCT	GCTTGCTCTC
	901	GCCTCTGCG	TGACCGCGAT	TTTCTCGCCC	CTCGCCTCCC	TCTGTCTGCC
	951	GGAAAACTAC	GCGCGCGTCC	GGTTTATCGT	CGTATCGTGT	ATGCTGCTCT
40	1001	CGCTGTTTTC	CACGCTGGTA	GA AATCAGCG	GCATCGGTTT	GAACGTCGTC
	1051	CGAAAAACAC	GCCCGATCGC	GCTCGCCACC	TTGGGCGCGC	TGGCGGCRAA
	1101	CTGCTGCTGT	CTGGGGCTTG	CCGTACCGTC	CGGCGCGCGG	CGGCGCGCGG
	1151	CGGTTGCTGT	TGCCGCTCA	TTTTGGCTGT	TTTTTGTGTT	CAAGACCGAA
	1201	AGCTCTGCG	GCCTGTGGCA	CGCGCTCAAA	CGCCTGCCGC	TTTATATGCA
	1251	CACATTTGTC	TGCTTGCGCT	CCTCGGCGGG	CTACACCTGC	TTGCGCACTC
45	1301	CGGCAAACTA	CCCCCTGTTT	GCGGCGGTAT	GGGCGGTATA	TCTGCGAGCG
	1351	TGCTCTCTGC	GCCACGGGAA	AGATTGTGAC	AAACTGTTTC	ATTATTGTA
	1401	AAAACAGGTT	TTCCCATAT	GA		

This encodes a protein having amino acid sequence <SEQ ID 378>:

50	1	MDTKRILGYA	AGSIGSAVLA	VIILPLLWY	FPADDIGRIV	IMQTAAGLTV
	51	SVLCLGLDQA	YVREYYAARD	KDTLFTKTLFL	PPLLAAAAA	ALLSRSPSL
	101	SEILFSLDDA	AAGIGLVLF	LSFLPIRFL	LVLRMGRAL	AFSSAQLVSK
	151	LAILLLLPLT	VGLLHFEPANT	AVLTAVAYALA	NLAAAPFLF	QNRCLKAVR
	201	RAPFSSAVLH	RGLRYGPIA	LSIIAYWGLA	SADRLFLKXY	AGLEQLGVYS
	251	MGISFGGAAL	LQSFISTVW	TPYIFRAIEA	NAPPARLARS	ABSAALLAS
55	301	ALCLTGI	FSP LASLLFPENY	AAVRFIVVSC	MLPPLFCTVL	EISGIGLN
	351	RKTRPIALAT	LGALAANLL	LGLAVPSSGA	RGA AVCAAS	FWLFFVKTE
	401	SSCRLWQPLK	RLPLYMHTLF	CLASSAAVTC	FGTANYPPLF	AGVWAVYLAG
	451	CILRHRKDLH	KLFHYLKKQG	FPL*		

ORF10a and ORF10-1 show 95.4% identity in 475 aa overlap:

60		10	20	30	40	50	60
orf10-1.pep		MDTKRILGYA	AGSIGSAVLA	VIILPLLWY	FPADDIGRIV	IMQTAAGLTV	SVLCLGLDQA
orf10a		MDTKRILGYA	AGSIGSAVLA	VIILPLLWY	FPADDIGRIV	IMQTAAGLTV	SVLCLGLDQA
65		10	20	30	40	50	60

		70	80	90	100	110	120
5	orf10-1.pep	YVREYXYADKDTL	FKFLFLP	PLLSAAIAALLSRP	LSLPSEIL	FSLD	DAAGLGLVFE
	orf10a	YVREYAAADKDTL	FKFLFLP	PLLSAAIAALLSRP	LSLPSEIL	FSLD	DAAGLGLVFE
		70	80	90	100	110	120
	orf10-1.pep	130	140	150	160	170	180
	orf10a	LSFLPIRIFLLVL	RMEGRALAFSSAQLV	PKLAILLXPLTVGLL	HFPA	NTAVLTAVYALA	
10		130	140	150	160	170	180
	orf10-1.pep	190	200	210	220	230	240
	orf10a	NLAAAFLFQNR	CLRKAVRHAPFSAV	LHRGXRYGIP	IALSSIA	YWGLSADRFL	FKKY
15		190	200	210	220	230	240
	orf10-1.pep	250	260	270	280	290	300
	orf10a	AGLEQLGVYS	MGISFGGALLFQ	SIFSTVTPYF	FR	AIENAPPARLSATAES	AALLAS
20		250	260	270	280	290	300
	orf10-1.pep	310	320	330	340	350	360
	orf10a	ALCXGIFSP	LASLLPENYA	AVRFIVWSCKP	PLFCTLAEISG	IGINVVAKTR	PIALAT
25		310	320	330	340	350	360
	orf10-1.pep	370	380	390	400	410	419
	orf10a	LGALAANLL	LGSLDRAV	PAR-PXGA	AVACASFWLFF	FAKTESSCR	WOPKRLPLYLHT
30		370	380	390	400	410	419
	orf10-1.pep	420	430	440	450	460	470
	orf10a	LFCLTSSAAYTC	FGTPN	ANYPLFAGV	WAAYLAGC	LIRHKLHLK	FLHYLKKGGFFFLX
35		420	430	440	450	460	470
	orf10-1.pep	480	490	500	510	520	530
	orf10a	LFCLASSAAYTC	FGTPN	ANYPLFAGV	WAAYLAGC	LIRHKLHLK	FLHYLKKGGFFFLX
40		480	490	500	510	520	530

ORF10 shows 94.1% identity over a 475aa overlap with a predicted ORF (ORF10.ng) from *N.*

orf10ng.pep	MDTKEILGYAAGSISGAVLAVILPLPLLSWYFPADDIGRIVLMQTAGLGVSVLCIGLDQA	60
orf10nm	MDTKEILXYAAGSISGAVLAVILPLPLLSWYFPADDIGRIVLMQTAGLGVSVLCIGLDQA	60
orf10ng.pep	YVREYYAAADKDTFKTLFLPPLFLSAIAALLSRFSPLSEILFSDIDAAAGISGLVILE	120
orf10nm	YVREYYAADKDTFKTLFLPPLFLSAIAALLSRFSPLSEILFSDIDAAAGISGLVILE	120
orf10ng.pep	LSFLPFRILLVLRMREGRALAFSAQLVFKLAILLXPLTVGLLHFFPANTAVTAVYALA	180
orf10nm	LSFLPFRILLVLRMREGRALAFSAQLVFKLAILLXPLTVGLLHFFPANTAVTAVYALA	180
orf10ng.pep	NLAAAAFLLFNQRCRLKAVRRAPFSPAVLHGRIGYRGLPALSSIAYWGLASADRFLFKY	240
orf10nm	NLAAAAFLLFNQRCRLKAVRHAPFSPAVLHGRXRYGIPIALSSIAYWGLASADRFLFKY	240
orf10ng.pep	AGLEQLGVYSMSISFGGAALLQISFSTVWTPYFRAIEENAPARLSATESAAGAILLAS	300
orf10nm	AGLEQLGVYSMSISFGGAALLQSFSTVWTPYFRAIEENAPARLSATESAAGAILLAS	300
orf10ng.pep	ALCETGISFPLASLLPENYAARVFTVSCMLPPLYFTAEITIGIGINVVVKRTRPIALAT	360
orf10nm	ALCXTGISFPLASLLPENYAARVFTVSCMLPPLYFTAEITIGIGINVVVKRTRPIALAT	360

-241-

		370	380	390	400	410
orf10ng.pep	LGALAA	NIIL	LLGL--AV	PSGGT	RGAAV	CAASF
5 orf10nm	LGALAA	NIIL	LLGLDRA	VPAR-EX	GAAV	CAASF
		370	380	390	400	410
		420	430	440	450	460
10 orf10ng.pep	LFCL	ASSA	YTCF	GT	PANY	PLFAG
orf10nm	LFCL	TSSA	YTCF	GT	PANY	PLFAG
		420	430	440	450	460

The complete length ORF10ng nucleotide sequence <SEQ ID 379> is:

1	ATGGACACAA	AAGAAATCCT	CGGCTACGCG	GCAGGCTCGA	TCGGCAGCGC
51	GGTTTTAGCC	GTCAATCATCC	TGCCGCTGCT	GTCGTGGTAT	TTCCcgcCGG
101	ACGACATCGG	GCGCATCGTG	CTGATGCAGA	CGCGCGCGGG	ACTGACGGTG
151	TCGGTATTGT	GCGTGGGCT	GGATCAGGCA	TACGTCCGCG	AATACTATGC
201	CGCGCGCGAC	AAAGACACTT	TGTTCAAAAC	CCTGTTCTGT	CGCGCGCTGC
251	TGTTTTCCGC	CGGATAGGCC	GCCTGCTGCT	TTTCCGCGCC	GTCCCTGCGG
301	TCTGAAATCC	TGTTTTGCTT	CGACGATGCC	CGCGCGCGCA	TCGGGCTGGT
351	GCTGTTGAA	CTGAGCTTCC	TGCCCATCCG	CTTCTCTTGA	CGGTGTTGCT
401	GATATGGAAG	GCGCGCCCTT	GCTCTTTCTGT	CGCGCAACT	CGTGCCTCAA
451	CTCGCCATTC	TGCTGCTGTT	CGCGCTGACG	GTCGGGCTGC	TGCATTTCC
501	GGCGAACACC	TCCGTCTGTA	CGCGCTTTTA	CGCGCTGGCA	AACCTTCCGG
551	CGCGCGCCTT	TTTGCTGTTT	CAAAACCGAT	GCGGCTGAA	GGCGCGCGGG
601	CGCGCGCGGT	TTTCCGCGCG	CGTCTGCAC	CGGGGCTGCA	GCTACGGCAT
651	ACCGCTGCAT	CTGAGCAGCC	TTGCTATTG	GGGGCTGGCA	TCGCGCGACC
701	GTTTGTTCTC	GAAAAATAT	GCGGGCCTGG	AACAGCTCGG	CGTTATTGCG
751	ATGGGTTATT	CGTTCGGCGG	GCGCGCATTA	TTGCTCCAAA	GCATCTTTTC
801	AACGGTCTGG	ACACCGTATA	TTTTTCGTGC	AATCGAAGAA	AACCGCAOCC
851	CGCGCGCGCT	CTCGGCAAGC	CGAGAAATCG	CGCGCGCGCT	GCTTGCTCTC
901	GCCTCTGCGC	TGACCGGAAAT	TTTCTGCGCC	CTCGGCTGCC	TGCTGCTGCC
951	GGAACATAC	CGCGCGCGCT	GCTTACCGT	CGTATCGTGT	ATGCTCGCGC
1001	CGCTGTTTGA	CACGCTGACC	GAAATCAGCG	GCAATCGTGT	GAACTCGCTC
1051	CGCAAAACGC	GTCCGATGCG	GCTTGCAACC	TGGGGCGCGC	TGGCGGCAAA
1101	CGTCTGCTGT	CTGGGGCTTG	CGTACCGTCT	CGCGCGCACG	CGCGCGCGGG
1151	CGGTTGCTGT	TGCCGCTGCA	TCTCTGTTGT	TTTTTGTGTT	CAAGACAGAA
1201	AGCTCCTGCG	GCGCTGGGCA	GCGCTCAAA	CGCTCGCGCG	TTTATATGCA
1251	CACATTGTTT	TGCCCTGGCT	CCTCGCGCGC	CTACACCTGC	TTGCGCACAC
1301	CGGCAAACTA	CCCCctgttt	gccgcgctat	GGGCGGCATA	TCTGCGCAGC
1351	TGCACTCTGC	GCCACCGGAA	AAATTTGCAC	AAACTGTTTC	ATTATTGGA
1401	AAAACAAGGT	TTCCATTAT	GA		

This encodes a protein having amino acid sequence <SEQ ID 380>:

1	MDTKEILGVA	AGSIGSAVIA	VIILPILLSWY	FPADDIGRIV	LMQTAAGLTV
51	SVLCLGLDQA	YVREYYAAAD	KDTLFKTLFL	FPLLFSAATA	AILLSRPSLP
101	SEILFSLDDA	AAGIGLVLFE	LSFLPIRELL	LIVRNEGRAL	AFSSAGLVPK
151	LAILLLEPLT	VGLLEFPANT	SVLTAVYALA	NLAAAFLFL	QNRCLKAVR
201	RAFFSPAVLH	RLRYGIVELA	LSLSAYWGLA	SADRFLKRY	AGLEGLGYG
251	MGISFGGAL	LLQSTFSTVW	TPYIFRAIEE	NATPARLSAT	AESAALLAS
301	ALCLTGIFSP	LASLLLPENY	AAVRFTVVSC	MLPPLFTYTL	ETSGIGLINV
351	RKTRIPALAT	LGALAAANLL	LGLAVPSGGT	RGAAVCAAS	FWLFFVKTE
401	SSCRWLQPLK	RLFLYMHYTL	CLASSAAYTC	FGTPANYPLF	AGVNAAYLAG
451	CILRRHKNLH	KLFHYLKKQG	FPL*		

ORF10ng and ORF10-1 show 96.4% identity in 473 aa overlap:

		10	20	30	40	50	60
orf10-1.pep	MDTKEILGVA	AGSIGSAVIA	VIILPILLSWY	FPADDIGRIV	LMQTAAGLTV	SVLCLGLDQA	
orf10ng-1	MDTKEILGVA	AGSIGSAVIA	VIILPILLSWY	FPADDIGRIV	LMQTAAGLTV	SVLCLGLDQA	
		10	20	30	40	50	60
		70	80	90	100	110	120
orf10-1.pep	YVREYYATAD	KDTLFKTLFL	PILLSAATA	AILLSRPSL	SEILFSLDDA	AGIGLVLFE	
orf10ng-1	YVREYYAAD	KDTLFKTLFL	PILLSAATA	AILLSRPSL	SEILFSLDDA	AGIGLVLFE	
		70	80	90	100	110	120

		130	140	150	160	170	180
5	orf10-1.pep	LSFLPTRFLLVLRMEGRALAFSSAQLVFKLAII	LLLP	TVGLLHFPANTAVLTAVYALA			
	orf10ng-1	LSFLPTRFLLVLRMEGRALAFSSAQLVFKLAII	LLLP	TVGLLHFPANTSVLTAVYALA			
		130	140	150	160	170	180
10	orf10-1.pep	NLAAAAFLLFONRCRLKAVRHAPFSPAVLHRLRGVLP	IALSS	SIAYWGLASADRLEFLKKY			
	orf10ng-1	NLAAAAFLLFONRCRLKAVRHAPFSPAVLHRLRGVLP	IALSS	SIAYWGLASADRLEFLKKY			
		190	200	210	220	230	240
15	orf10-1.pep	AGLEQLGVSMGIFSGGAALLQSIFSTVWTPYIFRAI	EENAPP	ARLSATAESAAALLAS			
	orf10ng-1	AGLEQLGVSMGIFSGGAALLQSIFSTVWTPYIFRAI	EENAPP	ARLSATAESAAALLAS			
		250	260	270	280	290	300
20	orf10-1.pep	ALCLTGIFSPLASLLLPENYAARVIVVSCMLPPLFY	TTIT	ISGIGLNVVRKTRPIALAT			
	orf10ng-1	ALCLTGIFSPLASLLLPENYAARVIVVSCMLPPLFY	TTIT	ISGIGLNVVRKTRPIALAT			
		310	320	330	340	350	360
25	orf10-1.pep	LGALAAANLLGLAVFSGGARGAAVACAASF	WFLFF	AFKTESSCRLWQPKRLPLYLHTLF			
	orf10ng-1	LGALAAANLLGLAVFSGGARGAAVACAASF	WFLFF	AFKTESSCRLWQPKRLPLYLHTLF			
		370	380	390	400	410	420
30	orf10-1.pep	CLTSSAAATCFGTTPANYPLFAGVWAAYLAGCIL	RHRKDL	HLKLFHYLKKGGFFLX			
	orf10ng-1	CLTSSAAATCFGTTPANYPLFAGVWAAYLAGCIL	RHRKDL	HLKLFHYLKKGGFFLX			
		430	440	450	460	470	
35	orf10-1.pep	CLTSSAAATCFGTTPANYPLFAGVWAAYLAGCIL	RHRKDL	HLKLFHYLKKGGFFLX			
	orf10ng-1	CLTSSAAATCFGTTPANYPLFAGVWAAYLAGCIL	RHRKDL	HLKLFHYLKKGGFFLX			
		430	440	450	460	470	

Based on this analysis, including the presence of a putative leader peptide and several transmembrane segments and the presence of a leucine-zipper motif (4 Leu residues spaced by 6 aa, shown in bold), it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 45

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 381>:

45	1..ATCCTGAAAC	CGCATAACCA	GCTTANGGAA	GACATCCAAC	CTGATCCGGC
51	CGATCAAAAC	GCCTTGTCCG	AACCGGATGC	TGCGACAGAG	GCAGAGCAGT
101	CGGATGCGGA	AAATGCTGCC	GACAAGCAGC	CGGTGCGCGA	TAAAGCCGAC
151	GAGGTTGAG	AAAAGCGCGG	CGAGCCGSA	CGGAGAGGCG	CGGACGGAC
201	GGCATGCT	AAGAAGCGC	TGACGAGGA	CGCTGACACA	ACCGTCAGGG
251	AAAAGCGCA	GAAGAAGAT	GCCGAACGG	TTAAATACGA	ACCGGTAAAA
301	CGCTCTAAG	AAACAGAGAA	AAAAGCTTCA	AAAGAAGAGA	AAAAGCGCGC
351	GAAGGAAAA	GTTGCACCCA	AACCAACCC	GGAACAAATC	CTCAACACGG
401	GCAgCATCGA	AAAGCGCGCG	AgTGCCCGCG	CCAAAGAAGT	GCAgAAAAATG
451	AA.AACGTCC	GACAAGCGCG	AAGC.AACGC	ATTATCTGCA	AATGGGCGCG
501	TATGCCGACC	GTCAAGCGCG	GGAAGGCGAG	CGTGCACAAAC	TGGCACTTCT
551	GGGCATATCT	TCCAAGGTGG	TGGTTTATCA	GCGCGGACAT	AAACCGCTTT
601	ACCGGGTGCA	AAGCGGCAAT	ATGTTCTCGC	ATGCGGTGA	

This corresponds to the amino acid sequence <SEQ ID 382; ORF65>:

60	1..ILKPHNQLKE	DIQDPADQN	ALSEPDAAE	AEQSDAENAA	DKQPVADKAD
51	EVEEKAGEPE	REEPDGAVR	KKALTEEREQ	TVREKAQKGD	AETVQIAQV

-243-

```

101 PSKETEKKAS KEEKKAKEK VAPKPTPEQI LNSGSIEXAR SAAAKEVQKM
151 XNVRCGGGSKR IICKWARMPT VRARKGSVFN WQSWAYLPRW SVIRRDIKRF
201 TGCKAAICLP MR*

```

Further work revealed the complete nucleotide sequence <SEQ ID 383>:

```

5      1 ATGTTTATGA ACAAATTTTC CCAATCCGGA AAAGTCTGT CCGGTTTTTT
      51 CTTCGGTTTG ATACTGGCGA CGGTCAATTAT TGCCGGTATT TTGTTTATC
101    101 TGAAACAGAG CGGTCAAAT GCGTTCAAAA TCCCGGTTTC GTCGAAGCAG
      151 CCTGCAGAAA CGGAATCTCT GAAACCGAAA AACCCAGCTA AGGAAGACAT
201    201 CCARCTTGAA CGGSCGATC AAAACGCCCT GTCCGAAAGC GATGCTGCGA
10      251 CAGAGCAGA CCACTCGGAT GCGGAAAAGT GTCCGACACA GCAGCCGCT
      301 GCGGATTAAG CCGACGAGGT TGAAGAAAAG GCGGGCGAGC CGGACCGGGA
351    351 AGAGCCGGAC GGACAGGCAG TGCCTAAGAA AGCGCTGACG GAAGAGCGTG
      401 AACAAACGCT CAGGGAAAAA GCGCAAGAA AGATGCCGA AACGGTTTAA
451    451 AAACAAGCGG TAACAACGCT TAAGAAGAAC GAGAAAAAAG CTTCAAAAGA
15      501 AGAGAAAAAG GCGGCGAAGG AAAAAGTTGC ACCCAACCA ACCCGGAAC
      551 AAATCTCTCA CAGCGGCAGC ATCGAAAAAG GCGGCGTGC GCAGCGCAAA
601    601 GAAGTGCAGA AAATGAAAAC GTCCGCAAG GCGGAAGCAA CGCATTATCT
651    651 GCAATGGGCG GCGTATGCGG ACCGTACAG GCGGAAAGG CAGCGTGCCA
20      701 AACTGGCAAT CTTGGGCATA TCTTCCAAG TGCTCGGTTA TCAGGCGGGA
      751 CATAAAAGCG TTTACCGGGT GCAAAGCGGC AATATGTCG CCGATGCGGT
801    801 GAAAAAATG CAGGACGAGT TGAAGAAAAC TGAAGTCGCC AGCCTGATCC
851    851 GTTCTATGGA AAGCAAAATA

```

This corresponds to the amino acid sequence <SEQ ID 384; ORF65-1>:

```

25      1 MFMNKFQSG KGLSGFFFL ILATVIAGI LFYLNQSGQN AFKIPASSKQ
      51 PAETELFKP NQPKEDIQPE PADQNALSEP DAATEAEQSD AEKADKQPV
101    101 ADKADVEEEK AGEPEEREED GQAVRKALAT EEREQTVREK AQKKDAETVX
      151 KQAVKPSKET EKKASKEEKK AAKEKVAPKP TPEQILNSGS IEKARSAAAK
201    201 EVQKMTSDK AEATHYLMQ AYADRQGAEG QRAKLAILGI SSKVVGQYAG
251    251 HKTLYRVQSG NMSDAVKKM QDELKKHEVA SLIRSIESTK*

```

30 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF65 shows 92.0% identity over a 150aa overlap with an ORF (ORF65a) from strain A of *N.*

meningitidis:

```

35      orf65.pep      10      20      30
      orf65a      ILKPHNQLKEDIQPDADQNALSEPDAAE
      30      40      50      60      70      80
      orf65a      ILAGILFYLNQSGNQAFKIPVPSKQPAETELKPKNQPKEDIQPEPADQNALSEPDAAE
40      40      50      60      70      80      90
      orf65.pep      AEQSDAENAAKQPVADKADVEEEKAGEPEREEDGQAVRKALTEEREQTVREKAQKID
      orf65a      AEQSDAEKAADKQPVADKADVEEEKAGEPEREEDGQAVRKALTEEREQTVREKAQKID
45      90      100      110      120      130      140
      orf65.pep      AETVKIQAVKPSKETEKKASKEEKKAAKEKVAPKPTPEQILNSGSIEXARSAAAKEVQKM
      orf65a      AETVKIQAVKPSKETEKKASKEEKKAAKEKVAPKPTPEQILNSGSIEXARSAAAKEVQKM
50      150      160      170      180      190      200
      orf65.pep      XNVRCGGGSKRIICKWARMPTVRARKGSVFNWQSWAYLPRWSVIRRDIKRF TGCKAAICLP
55      210      220      230      240      250      260
      orf65a      KTFDKAEATHYLMQAYADRRSAGGQRAKLAILGISKKVVGQAGHKTLYRVQSGNMSAD

```

The complete length ORF65a nucleotide sequence <SEQ ID 385> is:

```

1 ATGTTTATGA ACAAATTTTC CCAATCCGGA AAAGTCTGT CCGGTTTTTT
51 CTTCGGTTTG ATACTGGCGA CGGTCAATTAT TGCCGGTATT TTGTTTATC

```

101	TGACCCAGAG	CGGTCAAAAT	GGCTTCAAAA	TCCCGGTTC	GTGGAAGCAG
102	CTGTCCAGAT	CGGAATCCT	CGACCGGAAA	AACCGCGCTA	AGAGACACAT
151	CCGCGATG	CGGACCGT	GGAGACCG	GTGCGACCG	GTGCTCGGA
251	ACGAGCGAGA	CGCTCGAGT	GGAGACCG	CTCGCCAGC	CGACCGCGAG
301	CGCCACAAAC	CGCACAGGTT	TGCGCAAAAG	CGCGACGAGC	CGGACGGGGA
351	AAATGTGCAG	CGCACAGGCA	TGCGCAAAAG	AGCATCTACG	GAAGAGCGTG
401	ACAACACCGT	CGGGGAAJJA	CGCGCGAAGA	AGATGTCCGA	AACGCTTAA
451	AAACACGGG	TGAAGAACAT	GAAGAAACAT	CTTCAAAAGA	
501	AGAGAAAGAG	CGCGAGAGAG	AAAAGTTTCG	ACCCAAACCG	ACCCCGGAAC
551	AAATCTTCAA	CGGTACGAGC	CGCGCGAGAG	CGCTGCCAA	
601	AGATGTGCAG	ACATGAAAC	CGCCGACAGC	CGCGAGACGA	CGCATATTCT
651	GCAATTGGGC	CGGTATGCCG	ACCGCCGAGG	CGCGAGAGGG	CAGCGTCCCA
701	CACTGGCAAT	CTTGGGATCA	TCTTCCGAGG	TGGTGGTTTA	TCAGGCGGGA
751	CACTAAACCG	TTTACCGGGT	CGATATGCGG	ATATATGCTG	CGATATGCGT
801	GAATAAACAT	CAGGACAGAT	TGAATAAACA	TGAAGTCGCC	AGCCTGATCC
851	CTTCTATCGA	AAGCAAATTA			

This encodes a protein having amino acid sequence <SEO ID 386>:

```

1  MNFNKFSQSG KGLSGFFGLF ILATVIAGI LFVLNQGQGN AFKIPVPSKG
2  PAETILPKILP NOKPIDOPE PADQNALSEP DAAKEAQSDP AEKAADQKPV
20 101 AKADEAVEEEK ADEPEREKSD QGAQVKKATIL EEREQTIVGEK AQKKDKARSAK
151 KQAVPKSPKTE EKKASQKKK AEKQVAPKP PQTQLNGSE ILKEDASAAK
201 EVQKMTFPOK AEATHYLVOK AYADRSSAGE QRAKLAILGI SSKVVGYQAG
251 HKTLYRVOSG NMSADAVKMM ODELKHKVEA SLRISIESK*

```

ORF65a and ORF65-1 show 96.5% identity in 289 aa overlap:

25	orf65a.pep	MFMNKFQSQSGKLSGFFFGLLIATVILIAIGILFYLNQSSQNAFKIPVPSQPASTEILKKP	10	20	30	40	50	60
	orf65-1	MFMNKFQSQSGKLSGFFFGLLIATVILIAIGILFYLNQSSQNAFKIPASQPASTEILKKP	10	20	30	40	50	60
30	orf65a.pep	NQPKEDIQPEPADQNALSEPDAAEAEQSDAEKAAADKPQVADKADEVEEKADEPEREKSD	70	80	90	100	110	120
	orf65-1	NQPKEDIQPEPADQNALSEPDAAEAEQSDAEKAAADKPQVADKADEVEEKAGEPEREED	70	80	90	100	110	120
35	orf65a.pep	GQAVKKKALTEEREQTVTGSEKAKKKDAETVKKQAVKPSKETEKKASKEEKKAEKVAKP	130	140	150	160	170	180
	orf65-1	GQAVKKKALTEEREQTVREKAKKKDAETVKKQAVKPSKETEKKASKEEKKAEKVAKP	130	140	150	160	170	180
40	orf65a.pep	TPQIILNSGSIIEKARSAAREVQSMKTPDKAEATHYLMQGYADRRSABSGQRAKLAILGI	190	200	210	220	230	240
	orf65-1	TPQIILNSGSIIEKARSAAREVQSMKTPDKAEATHYLMQGYADRRSABSGQRAKLAILGI	190	200	210	220	230	240
45	orf65a.pep	SSKVGVYQAGHKTLRYVQSGNMSADAVKKMQDELKKHEVASLIRISIESKX	250	260	270	280	290	
	orf65-1	SSKVGVYQAGHKTLRYVQSGNMSADAVKKMQDELKKHEVASLIRISIESKX	250	260	270	280	290	

55 Homology with a predicted ORF from *N.gonorrhoeae*

ORF65 shows 89.6% identity over a 212aa overlap with a predicted ORF (ORF65.ng) from *N. gonorrhoeae*:

60

```
ORF65ng IIA GILL YL NQGG QNA FKIP ASKQPAET EIL KLN KPKED IQEPADQNALSEP DVAKK  
          ||| :|| |::|::|::|::|::|::|::|::|  
ORF65    ILKPNQLKE DIQFPDPAQNALSEPDAAE  
                                10      20      30
```

-245-

		90	100	110	120	130	140
ORF65ng	AEQSDAEKADKQPVADKADAEVEEKAGEPEREEDPGQAVRKALTEEREQTVREKAQKKD						
5	ORF65	AEQSDAENAAADKQPVADKADAEVEEKAGEPEREEDPGQAVRKALTEEREQTVREKAQKKD					
		40	50	60	70	80	90
		150	160	170	180	190	200
ORF65ng	AETVKKKAVKPSKETETKASKKEEKKAAEKVAPKPTPEQILNSRSIEKARSAAAKEVQKM						
10	ORF65	AETVKIQKAVKPSKETETKASKKEEKKAAEKVAPKPTPEQILNSGSEKARSAAAKEVQKM					
		100	110	120	130	140	150
		210	220	230	240	250	260
ORF65ng	KNFGQGGSRRIICKWARMNPNGARKGSPVNWQSWAYLPKWSAIRRDIKRFCTACKAAICPP						
15	ORF65	XNVRRGGSRRIICKWARMNPVRRKGSVPNWQSWAYLPFRWSVIRRDIKRFCTCKAAICLP					
		160	170	180	190	200	210
	ORF65ng	MR					
		II					
20	ORF65	MR					

An ORF65ng nucleotide sequence <SEQ ID 387> was predicted to encode a protein having amino acid sequence <SEQ ID 388>:

	1	MFNNKFSQSG	KGLSGFFFL	ILATVIIAGI	LLYLNQGGGN	AFKIPAPSKQ
25	51	PAETETILKIK	NQPKEDTQPE	PADQNALSEP	DVAKEAQSD	AEKADKQPV
	101	ADKADAEVEEK	AGEPEREED	GQAVRKALTE	EEREQTVREK	AQKDAETVK
	151	KKAVKPSKET	EKKASKKEKK	AAKEKVAPKP	TPEQILNSRS	IEKARSAAAK
	201	EVQKMKNFQ	GGSRRIICKW	ARMNPNGARK	GSPVNWQSWA	YLPKWSAIRR
	251	DIKRFCTACKA	AIQPPMR*			

30 After further analysis, the complete gonococcal DNA sequence <SEQ ID 389> was found to be:

	1	ATGTTTATGA	ACAAATTTTC	CCAAATCCGGA	AAAGGTCTGT	CGGTTTCTTT
	51	CTTCGGTTTG	ATACTGSCAA	CGGTCAATTAT	TGCCGGTATT	TTGCTTTTAT
	101	TGAACCAAGG	CGGTCAAAAT	GGTTCACAAA	TCCCGGTCTC	GTCGAAGCAG
35	151	CCTGCAGAAA	CGGAATTCCT	GAACACTGAAA	AACCCAGCCTA	AGGAAGACAT
	201	CCAACTTGAA	CCGGCCGATC	AAAAACGCTT	GTCGGAACCG	GATGTTGCGA
	251	AAGAGGCAGA	GCAGTCGAT	CGGGAARAAG	CTGCCGACAA	GCAGCCCGTT
	301	CGCGACAAAG	ccgacgAGGT	TGARAAGAAg	ccgacgAGcg	cggAACGGga
	351	aGAGCGGCAC	ggACAGGCGAG	TGCGCAAGAA	AGCATCTAGg	gAAGAgcGTG
	401	AACAAACGgt	cagggAAAAA	GCGCagaaga	AAGATGCCGA	AACGgTTAAA
40	451	AAcacaGCGg	tAaaaccgcgc	tAAAGAAACA	gagaaaaaag	cTtcaaaaaa
	501	agagaaaaaag	gcggcgaaag	aaaAAGgtgc	accacaaacc	accccggaac
	551	aaatcctcaa	cagccgcGagc	atcgaaaaaag	cgctgtagtc	cgctgcacaa
	601	gaAggtcaga	AAatgaaaaa	ctTtggcgaa	gcgcGaagcc	aacgattaAT
	651	CTGcaaatgg	gcgcgtatgc	cgaccggtccg	gagccggaga	gggcagcggt
45	701	ccaaAActg	gcAActgtgc	atatctTccg	aagtgtcgGg	CTATCAGGCG
	751	GGACATAAAA	GCCTTTACCG	CGTGCAAAgc	GGCAatatgt	ccgcgatgac
	801	gGTGAAAAAA	ATGCAGGACG	AGTTGAAAAA	GCATGGGGgt	gcCAGCCTGA
	851	TCCGTGcgAT	TGAAGGCAAA	TAA		

This encodes the following amino acid sequence <SEQ ID 390>:

	1	MFNNKFSQSG	KGLSGFFFL	ILATVIIAGI	LLYLNQGGGN	AFKIPAPSKQ
50	51	PAETETILKIK	NQPKEDTQPE	PADQNALSEP	DVAKEAQSD	AEKADKQPV
	101	ADKADAEVEEK	AGEPEREED	GQAVRKALTE	EEREQTVREK	AQKDAETVK
	151	KQAVKPSKET	EKKASKKEKK	AAKEKVAPKP	TPEQILNSRS	IEKARSAAAK
	201	EVQKMKNFQ	GGSRRIICKW	ARMPTVRSAE	GORAKLAILG	ISSEVVGYYA
55	251	GKHTLYRVQS	GNMSADAVKK	MQDELKHHGV	ASLIRALIEGK	*

ORF65ng-1 and ORF65-1 show 89.0% identity in 290 aa overlap:

		10	20	30	40	50	60
orf65-1.pep	MFNNKFSQSGKGLSGFFFLILATVIIAGILFYLNQGGNAFKIPASSKQPAETETILKPK						
60	orf65ng-1	MFNNKFSQSGKGLSGFFFLILATVIIAGILFYLNQGGNAFKIPASSKQPAETETILKPK					
		10	20	30	40	50	60

-246-

		70	80	90	100	110	120
	orf65-1.pep	NQPKEDIQPEPADQNALSEPDAATEAEQSDAEKAAADKQPVADKADVEEKAAGEPEREEDP					
5	orf65ng-1	NQPKEDIQPEPADQNALSEPDAATEAEQSDAEKAAADKQPVADKADVEEKAAGEPEREEDP	70	80	90	100	110
							120
	orf65-1.pep	GQAVRKKALTEEREQTVREKKAQKKDAETVKKQAVKPSKETTEKKASKEEKKAAKEKVAEKP	130	140	150	160	170
10	orf65ng-1	GQAVRKKALTEEREQTVREKKAQKKDAETVKKQAVKPSKETTEKKASKEEKKAAKEKVAEKP					
			130	140	150	160	170
							180
	orf65-1.pep	TPEQILNSGSIIEKARSAAAKEVQKMKTSKAEATHYL-QMGAYADRSQAEQGQRAKLAITLG	190	200	210	220	230
15	orf65ng-1	TPEQILNSRSLIEKARSAAAKEVQKMKTFGQGGSQRIICKWARMPTVRSQAEQGQRAKLAITLG					
			190	200	210	220	230
							240
	orf65-1.pep	ISSKVVGVAQGHKTLTYRVQSGNMSADAVKKMQDELKKHEVASLIRISIESKX	240	250	260	270	280
20	orf65ng-1	ISSEVVGVAQGHKTLTYRVQSGNMSADAVKKMQDELKKHGVASLIRAEIGKX					
			240	250	260	270	280
							290

- 25 On this basis, including the presence of a putative transmembrane domain in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 46

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID

- 30 391>:

	1	ATGAACCAAG	ACATCACTTT	CCTCACCCCTG	TTCTACTCG	GTxTCTTCGG
	51	CGGAACGCAC	TGCATCGGTA	TGTGCGGCGG	ATTAAGCAGC	GcGTTTGS.s
	101	TCCAACTCCC	CCGCGATATC	AACCGCTTTT	GGCTGATCCT	GCTGCTTAAC
	151	ACAGGACGGG	TAAGCAGCTA	TACGGCAATC	GGCCTGATAC	TCGGATTAAT
35	201	CGGACAGGTC	GGGGTTTCAC	TCGACCAaAC	CCGCGTCTCT	CAGAATATTT
	251	TATACACGGC	CGCCAACTCT	CTGCTGCTCT	TTTTAGGCTT	ATACTTGAGC
	301	GGTATTCTTT	CCTTGGCGGC	AAAAATCGAG	AAATCGCA	AACCGATATG
	351	GGGAAACCTG	AACCGATATC	TCAACCGCT	GTTCACCAT	AAATCGATAT
	401	CCGCTGCTCT	TGCGTCGGA	ATATTATGG	GTCGCTGCC	GTCGGAATG
40	451	GTTTACACGG	CGTCCGCTTA	CGCGCTGGGA	AgCGCTAGT	CGGCAACGGG
	501	CGGGTTATAT	ATGCTTGCTT	TGCACTGGG	TACGCTGCC	AATCTTATAG
	551	CAATCGGCAT	TTTTCCCTG	CAACTGAaA	AAATCATGCA	AAACCGATAT
	601	ATCCGCTGCT	GTACGGGATT	ATCCGATAT	TTATGGGCAT	TATGGAATAT
	651	TGCGCTGCTG	TGCTGTGTA			

- 45 This corresponds to the amino acid sequence <SEQ ID 392; ORF103>:

	1	MNHDITFLTL	FLLGXFGGTH	CIGMCGGLSS	AFXXQLPPhi	NRFWLILLLN
	51	TGRVSSYTAI	GLILGLIGQV	GVSLDQTRVL	QNILYTAANL	LLFLGLLYLS
	101	GISSLAAKIE	KIKKPIWRNL	NPILNRLLPI	KSIPACLAAGV	ILWGLPFGSL
	151	VYSASLYALG	SGSAATGGLY	MLAFALGTLF	NLLAIGIFSL	QLXKIMNRY
50	201	IRLCTGLSVS	LWALNKLAVL	WL*		

Further work elaborated the DNA sequence <SEQ ID 393> as:

	1	ATGAACCAAG	ACATCACTTT	CCTCACCCCTG	TTCTACTCG	GTITCTTCGG
	51	CGGAACGCAC	TGCATCGGTA	TGTGCGGCGG	ATTAAGCAGC	GGCTTTCGCG
	101	TCCAACTCCC	CCGCGATATC	AACCGCTTTT	GGCTGATCCT	GCTGCTTAAC
55	151	ACAGGACGGG	TAAGCAGCTA	TACGGCAATC	GGCCTGATAC	TCGGATTAAT
	201	CGGACAGGTC	GGCGTTTCAC	TCGACCAaAC	CCGCGTCTCT	CAGAATATTT
	251	TATACACGGC	CGCCAACTCT	CTGCTGCTCT	TTTTAGGCTT	ATACTTGAGC
	301	GGTATTCTTT	CCTTGGCGGC	AAAAATCGAG	AAATCGCA	AACCGATATG

351 GGGAAACCTG AACCGGATAC TCAACCGGCT GTTACCCATA AAATCCATAC
 401 CGGCTGCCT TCGCGTCGGA ATATTATGGG GCTGGCTGCC GTGCGGACTG
 451 GTTTACAGCG CGTGCCTTTA CGCGCTGGGA AGCGGTAGTG CCGCAACGGC
 501 CGGGTTATAT ATGCTTGCCT TTGCACTGGG TACGCTGCC AATCTTTAG
 551 CAATCGGCAT TTTTCCCTG CAACGNAAAA AAATCATGCA AAACCGATAT
 601 ATCCGCGCTG GTACGGGATT ATCCGTAATCA TTATGGGCAT TATCGAAACT
 651 TGGCGTCTG TGGCTGTA

This corresponds to the amino acid sequence <SEQ ID 394; ORF103-1>:

1 MNHDTFLTL FLLGFFGGTH CIGMCGGLSS AFALQLPPIH NRFWILLLN
 10 TGRVSSYTAI GLILGLIGQV GVSLDQTRVL QNIIYTAANL LLLFLGLYLS
 101 GISSLAAKIE KIGKPIWRNL NFILNRLPI KSIAPCLAVG ILWGWLPCGL
 151 VYSASLYALG SGSAAATGGLY MLAFALGTLP NLLAIGIFSL QKKIMONRY
 201 IRLCTGLSVS LNALWKLAVL WL*

Computer analysis of this amino acid sequence gave the following results:

15 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF103 shows 93.8% identity over a 222aa overlap with an ORF (ORF103a) from strain A of *N.*

meningitidis:

		10	20	30	40	50	60
20	orf103.pep	MNHDTFLTLFLLGFFGGTHCIGMCGGLSSAFALQLPPIHNRFWILLNLTGRVSSYTAI					
	orf103a	MNXDITFLTLFLLGFFGGTHCIGMCGGLSSAFALQLPPIHNRXWILLNLTGRVSSYTAI					
		10	20	30	40	50	60
25	orf103.pep	GLILGLIGQVGSVDQTRVLQNIIYTAANL LLLFLGLYLSGISSLAAKIEKIGKPIWRNL					
	orf103a	GLILGLIGQVGSVDQTRVXQNIIYTAANL LLLFLGLYLSGISSLAAKIEKIGKPIWRNL					
		70	80	90	100	110	120
30	orf103.pep	NFILNRLPIKSIAPCLAVGILWGWLPCGLVYSASLYALGSGSAAATGGLYMLAFALGTLP					
	orf103a	NFILNRLPIKSIAPCLAVGILWGWLPCGLVYSASLYALGSGSAAATGGLYMLAFALGTLP					
		130	140	150	160	170	180
35	orf103.pep	NLLAIGIFSLQKKIMONRYIRLCTGLSVSLNALWKLAVLWLX					
	orf103a	NLXAIGIFSLQKKIMONRYIRLCTGLSVSLNALWKLAVLWLX					
		190	200	210	220		
40	orf103a	NLXAIGIFSLQKKIMONRYIRLCTGLSVSLNALWKLAVLWLX					
		190	200	210	220		

The complete length ORF103a nucleotide sequence <SEQ ID 395> is:

1 ATGACCCAG ACATCACTTT CTTCACTCTG TTCTACTGCG GTTTCITCGG
 45 CGGAAACGAC TCGATCGGTA TTCTCGGCGC GCGTTTGGCC
 101 TCCAACTCCC CCGCATATC AACCGCTTNT GCGCTAGCTC GCTGCTTAAC
 151 ACAGGAAGGG TAAGCAGCTA TACGCGAATC GCGCTAGATC TCGGATTAT
 201 CGGACAGGTC GCGCTTTTCA TCGACCAATC CCGCGTCNNG CAGAATATTT
 251 TATACACGGC CGCAACCTC CTGCTGCTCT TTTTAGGCTT ATACTTGAGC
 301 GGTATTTCCT CTTGGCGGCG AAAAATCGAG AAAATCGGCA AACCGATATC
 351 CGGGAACCTG AACCGGATAC TCAACCGGCT GTTACCCATA AAATCCATAC
 50 CGGCTGCCT TCGCGTCGGA ATATTATGGG GCTGGCTGCC GTGCGGACTA
 451 GTTTACAGCG CGTGCCTTTA CGCGCTGGGA AGCGGTAGTG CCGCAACGGC
 501 CGGGTTATAT ATGCTTGCCT TTGCACTGGG TACGCTGCC AATCTTTAG
 551 CAATCGGCAT TTTTCCCTG CAACGNAAAA AAATCATGCA AAACCGATAT
 601 ATCCGCGCTG GTACGGGATT ATCCGTAATCA TTATGGGCAT TATCGAAACT
 55 TGGCGTCTG TGGCTGTA

This encodes a protein having amino acid sequence <SEQ ID 396>:

1 MNHDTFLTL FLLGFFGGTH CIGMCGGLSS AFALQLPPIH NRXWILLLN
 51 TGRVSSYTAI GLILGLIGQV GVSLDQTRVX QNIIYTAANL LLLFLGLYLS
 101 GISSLAAKIE KIGKPIWRNL NFILNRLPI KSIAPCLAVG ILWGWLPCGL

151 VYSASLYALG SGSAATGGLY MIAFALGTLP NLXAGIFSL QLXKIMQNR
 201 IRLCTGLSVS LMAWKLAVL WL*

ORF103a and ORF103-1 show 97.7% identity in 222 aa overlap:

5	orf103a.pep	MNXDITFLTLFLGFFGGTHCIGMCGGLSSAFALQLPPHINRXWLLILLNTGRVSSYTAI	10	20	30	40	50	60
	orf103-1	MNHDITFLTLFLGFFGGTHCIGMCGGLSSAFALQLPPHINRFWLLILLNTGRVSSYTAI	10	20	30	40	50	60
10	orf103a.pep	GLILGLIGQGVSLDQTRVQNILYTAANLLFLGLYLSGSISSLAKEIKIGKPIWRNL	70	80	90	100	110	120
	orf103-1	GLILGLIGQGVSLDQTRVQNILYTAANLLFLGLYLSGSISSLAKEIKIGKPIWRNL	70	80	90	100	110	120
15	orf103a.pep	NPILNRLLEPKSIPACLAAGVILGWLPGLVYSASLYALGSGSAATGGLYMIAFALGTLP	130	140	150	160	170	180
	orf103-1	NPILNRLLEPKSIPACLAAGVILGWLPGLVYSASLYALGSGSAATGGLYMIAFALGTLP	130	140	150	160	170	180
20	orf103a.pep	NLXAGIFSLQLXKIMQNRIRLCTGLSVSLWALWKLAVLWL	190	200	210	220		
	orf103-1	NLLAIGIFSLQLXKIMQNRIRLCTGLSVSLWALWKLAVLWL	190	200	210	220		

Homology with a predicted ORF from *N. gonorrhoeae*

ORF103 shows 95.5% identity over a 222aa overlap with a predicted ORF (ORF103.ng) from *N.*

30 *gonorrhoeae*:

orf103.pep	MNHDITFLTLFLGFFGGTHCIGMCGGLSSAFYXQLPPHINRFWLLILLNTGRVSSYTAI	60
orf103ng	MNHDITFLTLFLGFFGGTHCIGMCGGLSSAFALQLPPHINRFWLLILLNTGRVSSYTAI	60
35 orf103.pep	GLILGLIGQGVSLDQTRVQNILYTAANLLFLGLYLSGSISSLAKEIKIGKPIWRNL	120
orf103ng	GLILGLIGQGVSLDQTRVQNILYTAANLLFLGLYLSGSISSLAKEIKIGKPIWRNL	120
40 orf103.pep	NPILNRLLEPKSIPACLAAGVILGWLPGLVYSASLYALGSGSAATGGLYMIAFALGTLP	180
orf103ng	NPILNRLLEPKSIPACLAAGVILGWLPGLVYSASLYALGSGSAATGGLYMIAFALGTLP	180
orf103.pep	NLLAIGIFSLQLXKIMQNRIRLCTGLSVSLWALWKLAVLWL	222
45 orf103ng	NLLAIGIFSLQLXKIMQNRIRLCTGLSVSLWALWKLAVLWL	222

The complete length ORF103ng nucleotide sequence <SEQ ID 397> is:

1	ATGAACACAG	ACATCACTTT	CCTCACCCCTG	TTCTGCTCG	GTTTCTTGG
51	CGGAACCTAC	TGCATCGGTA	TGTGCGGCGG	ATTAAGCAGC	GCCTTTGGGC
101	TCCAACCTCCC	CCGCGATATC	AACCGCTTTT	GGCTGATTCT	GCTGCTTAAC
151	ACAGGACGGA	TAAGCAGCTA	TACGGCAATC	GGCCTGATGC	TGGGATTAAT
201	CGGACAACTC	GSCATTTTAC	TCGACCAAA	ccgcgTCTGT	CAAAATATTT
251	tatacacagc	ctccaaCCTC	CTGCTGCTCT	TTTTAGGCTT	ATACTTGAGC
301	GGTATTTCCT	CTTGGCGGCG	AAAAATCGAG	AAAAATCGCA	AACCGATATG
351	GCGCAACCTG	AACCGGATAC	TCAACCGGCT	GCTGCCCATTA	AAATCCATAC
401	CCGCTGCTCT	TGCTGTCGGA	ATATTATGGG	GCTGCTGCC	GTGCGGACTG
451	GTTTACAGCG	CATCACTTTA	CGCCTGGGGA	AGCGGTAGTG	CGACACCGG
501	CGGACTGTAT	ATGCTTGCTT	TTGCACTGGG	TAGCTGCC	AAATCTTGG
551	CAATCGGCAT	TTTTTCCCTG	CAACTGAAAA	AAATCATGCA	AAACCGATAT
601	ATCCGCGCTGT	GTACAGGATT	ATCCGTATCA	TTATGGGCAT	TATGGAAGCT
651	TGCCGTCCTG	TGGCTGTAA			

This encodes a protein having amino acid sequence <SEQ ID 398>:

1 MNHDTFTLTL FLLGFFGGTH CIGMCGGLSS AFALQLPPHI NRFWLILLIN
 51 TGRISSTYAI GLMLGLIGQL GISLQDTRVL QNILYASNL LLLFLGLYLS
 101 GISSLAARKE KIGKPIWRNL NEILNRLPI KSPACLAVG ILWGLPGL
 151 VYSASLYALG SGSATTGGY MLAFALSTLP NLLAIGIFSL QLKRMONRY
 201 IRLCTGLSVS LMAWKLAVL WL*

In addition, ORF103ng and ORF103-1 show 97.3% identity in 222 aa overlap:

		10	20	30	40	50	60
orf103-1.pep		MNHDTFTLTLFLLGFFGGTHCIGMCGGLSSAFALQLPPHINR					
orf103ng		MNHDTFTLTLFLLGFFGGTHCIGMCGGLSSAFALQLPPHINR					
		10	20	30	40	50	60
orf103-1.pep		GLILGLIGQVGSLSLQDTRVLQNILYTAANLLFLGLYLSG					
orf103ng		GLILGLIGQLGISLQDTRVLQNILYTAANLLFLGLYLSG					
		70	80	90	100	110	120
orf103-1.pep		NPILNRLPIKSPACLAUGILNGLWGLVYSASLYALGSGSAATGGLYMLAFALSTLP					
orf103ng		NPILNRLPIKSPACLAUGILNGLWGLVYSASLYALGSGSAATGGLYMLAFALSTLP					
		130	140	150	160	170	180
orf103-1.pep		NLLAIGIFSLQLKRMONRYIRLCTGLSVSLMAWKLAVLNLX					
orf103ng		NLLAIGIFSLQLKRMONRYIRLCTGLSVSLMAWKLAVLNLX					
		190	200	210	220		
orf103-1.pep							
orf103ng							

Based on this analysis, including the presence of a putative leader sequence (double-underlined) and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 47

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 399>:

1	ATGGAAAC	AAAGCCGCT	CCTAGCCTT	CGCTGGCAC	TTTGGCGGC
51	GATGACGTGG	GGAACGCTGC	CGAT.TCCGT	GGCGCAGTA	TTGAAGTTTG
101	TCGATGCGCC	GAGCGTGGTG	TGGGTGGCTT	TTACCGTGGC	GGCGCGGTA
151	TTGTTTGT	TGCTGGCAT	GGCGGGCGG	CTGCcGAAGC	GGCGcGATT
201	TTTCTGGTG	CTCATTCCAG	CTGCTGCTGC	TGCGCGGGC	GGCGATTTCG
251	GCAACCTTG	TGCTGATTGC	CCAAGGGCTG	CATTATATTT	CGCGACCCAC
301	GACGCAAGT	TTGTGGCAGA	TTTCCCGTGT	TACAGTATT	GTWGTGGTG
351	TGTGGTGT	TAAAGACCGG	ATGACTGCGC	CTCAGAAAA	CGGCTTGGTT
401	TTGCTGCTG	CCGGTTTGCT	TATGATATTT	AACGATAAAT	TCGGCGAGTT
451	GTCGGGTTG	GGCGGATATG	C.AAGGGCGT	GTTGCTGTGT	GGCGCAGGCA
501	GTATGGCATG	GGTGTGTAAT	GGCGTGGCGC	AAAAGCTGCT	GTGCGGCGCA
551	TTCCGGCCCG	AACGATTCT	GCTGTTGATT	TATGGCGCAA	GTGCGCGCGT
601	GTTCCTGCG	TTTGGCGAAC	CGGACACAT	CGGAAGTATG	GACGGTAGT
651	TGGCGTGGT	ATGATATTGG	TATTGCTGCT	TGAATACGTT	ATCGGTAC
701	GGCTCGTTG	GGGAGGCGTT	GAAACATTGG	GAGGCTTCCA	AAGTCAGCGC
751	GGTAACAAC	TTGCTCCCGG	TGTTTACCGT	AATAAATACT	TGCTCGGGC
801	ATTATGTGAT	GCCTGAAACT	TTTGGCGCGC	CGGA..	

This corresponds to the amino acid sequence <SEQ ID 400; ORF104>:

1	MENQRLLGF	RLALLAAMTW	GTLFXSVRQV	LKFVDAPTLV	WVRFTVAADV
51	LFVLLALGGR	LPKRDFSWC	SERLLLLGVA	GISANFLVIA	QGLHYISPTT
101	TOVLQGISPF	TMIVGVVLVF	KDRMTAAQKI	GLVLLLAGLL	MYFNDFKFGEL
151	SLGLAYKGV	LLCAAGSMAN	VONAVAKLL	SAQFGPQIL	LLIYAASAV
201	FLFPAEPAHI	GSMDDGLTAW	CIAYCLNLT	IGYGSFGEAL	KHWEASKVSA

-250-

251 VTTLLPVETV INTLLGHYVM PETFAAP...

Further work revealed further partial DNA sequence <SEO ID 401>:

	1	ATGTGAAACG	GAAGCCCGCT	CTCAGGCTGT	GGCTTGGCAC	TTTGGCGCG
5	51	GATGACGTGG	GAAAGCGCTGC	GATGTCGGCT	GGCGGACGAT	TTGAAGTTTG
	101	TCGATCGGCC	GACGCGTGTG	TGTACGCTGTT	TTACCGCTGT	GGCGCGCGTA
	151	TTGTTGTTTGT	TGCTTGGCATC	GGGCGGCGGAC	CTGCGGAGAC	GGCGGAGATT
	201	TTCTTGGTCG	TGCTCATCGGC	TGCTGCTGCT	GGCGTGGCG	GGCATTTCGG
	251	CNACTTTTGT	GCTGATTGCC	CAGGGGCTGC	ATATATATT	GGCGACACGG
10	301	ACCGAGTTT	TGTGGGACAT	TGCGCCGTT	ACGATGATT	TGTCGGTGT
	351	GTGCTGTGTT	AAAGACCCGA	TGACTGGCGC	TACAGAAATC	GGCTTGCTTT
	401	TGCTGCTTTC	CGGTTGCTCT	ATGTTTTTTA	ACGATAAAAT	CGCGAGATTG
	451	TCGGGTTTGG	GCGCTATGCG	ATGGTCTGTC	TGTCGTGTGT	CGCGAGGCGT
	501	TATGGCATAT	GTGCTTATAT	CGGTGGCGCA	AAAGCTGCTG	TGCGCGCAAT
15	551	TGCGGCGCGA	ACGATGTTCT	GCTGTGATT	ATGCGGCAAG	GGCGCGGTTT
	601	TTCTCGGCTG	TGCGGACATC	GGGCGACATC	AGGAGTTTGT	ACCGTAGCTT
	651	GGCTGGTGGT	ATGTCAGCTT	ATGTCAGCTT	ATGTCAGCTT	ATGTCAGCTT
	701	GCTGCTTCGG	CGAGGCGCTT	AAACATCTAA	AGGCTTCCA	AGTCACGAGC
	751	GTGACAACT	TGCTCCCGCT	GTTTACCGTA	ATATMHTCT	TGCTCGGGCA
	801	TTATGTTGAT	CTCGGAACTT	TTGCCGCGGAT	GGG	

20 This corresponds to the amino acid sequence <SEO ID 402: ORF104-1>:

25

1	MEMORPELGG	ALALLAAALTA	GTLPATVAVQG	LKFVDAPTLV	WGLTPTAAAV
5	LFVLLAALGG	LKRRDFDWC	SRFRLLLGVA	GISANFVLTA	QVRHYISPTF
101	TVQLVQISPV	LTIVVGSMLV	KRMTAAQKI	GLVLLIAGLL	MFNFNDKFGEL
151	SGLGAYAKGF	LLCAGSGMAW	VCYVAAQKLL	SAQFGPQOIL	LLIYAAASAA
201	LPFFAEPAHT	GLSDGTLAW	CFAYCNCNLT	IGYSGPGEAL	KHWEASKVSA
251	VTTLFVGVTV	TXSLGLGVVW	PETFAPE...		

Computer analysis of this amino acid sequence gave the following results:

Homology with hypothetical HI0878 protein of *H. influenzae* (accession number U32769)

ORF104 and HI0878 show 40% aa identity in 277aa overlap:

30	orf104	4	QRLPLGFLIALAMWGTGLTPXSVVQVLKVDAPTLVWXXXXXXXXXXXXXXXXXXXXP- 62
			Q-PLLGF AL+ AM WG+LP +++QVL ++A T+VW 62
	HI0878	3	QQLPLGFTFALITAMAWGSLPIALQVLSVMNAQTIVWYRFIAAVSLALLLAYKKQLPE 62
35	orf104	6	--RRRDSWCFSRLLLLGVAGISANFVLIAQGLYISPTTQVLMQISFPTMVGVVLVF 120
			K R+ ++W +--+L+VG ++N+L+ + L+YI P+ Q+ +S F M++ GVL+F 120
	HI0878	6	LMKVQYAW----IMLIGVIGLTNFFLLSSSLNLTPEVSQAQIFHLSSFGMLICGVLF 118
	orf104	12	KDRMTAAQKIXXXXXXXXXXXMFKGDSGLGAYXKGVLLCAAGSMCAINAVAKLL 180
			K+++ QKI +---FND+F +GL Y GV+L G++ WV +AKL+ 180
40	HI0878	119	KEKLGHQKIGLFLLLIGLGFNFNDRDFAFGLAQYTSGVILGVGGALTIWAYGMAQKL 178
	orf104	181	SAQFGPQQILLIYAASAAVFLFPAEPAHIGSMOGLTLAGVCIAYCCNLILIGYSGEAL 240
			F QOILL++Y F A+ A+ + LA +C YCCNLITIGYSGE+ EAL 240
	HI0878	179	LRKFSQQIILLMLGCAIAIRNMFADFSQVQLT-PALICFTYCCNLITIGYGYAEL 237
45	orf104	241	KHNWASKVSATVTLPTVITVNTLIGVYQVMTFAAP 277
			W+ SKVS V TL+P+T++ + + HY P FAAP 277
	HI0878	238	KRWDSKVSQVTLTVLPELTITLFSHAIHYFSPADFAAP 274

50 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF104 shows 95.3% identity over a 277aa overlap with an ORF (ORF104a) from strain A of *N.*

meningitidis:

55

orf104.pep	10	20	30	40	50	60
	MENQRPLGLGFRIALLAAMTWTGLPSVRQVKLFVDAPTLVWVRFTVAAVLFLVLLAGGR					
orf104a	10	20	30	40	50	60
	MENQRPLGLGFALLAAMTWTGLPIAVRQVKLFVDAPTLVWVRFTVAAVLFLVLLAGGR					
	70	80	90	100	110	120

-251-

orf104.pep	LPKRRD	FSWCSF	RLLLG	VAGISAN	FVLIAQ	GLHYIS	PTTTQ	VLWQIS	PFMTI	VVGVLV	F
orf104a	LPKRRD	FSWCSF	RLLLG	VAGISAN	FVLIAQ	GLHYIS	PTTTQ	VLWQIS	PFMTI	VVGVLV	F
		70	80	90	100	110	120				
orf104.pep	KDRMTA	AQKIGL	VLLLAG	LMFYFN	DKFGE	LSGLG	AYAKG	VLLCA	AGSM	AWCN	VAQKLL
orf104a	KDRMTA	AQKIGL	VLLLAG	LMFYFN	DKFGE	LSGLG	AYAKG	VLLCA	AGSM	AWCN	VAQKLL
		130	140	150	160	170	180				
orf104.pep	SAQFGP	QQIILL	LIYAAS	AAVFLP	FAEPA	HIGSM	DGTLA	WVCIA	YCCLN	TLIGY	GSFG
orf104a	SAQFGP	QQIILL	LIYAAS	AAVFLP	FAEPA	HIGSM	DGTLA	WVCIA	YCCLN	TLIGY	GSFG
		190	200	210	220	230	240				
orf104.pep	KHWEAS	KVSAV	TTLFV	PTVIN	TLGH	YVMP	ETFA	AP			
orf104a	KHWEAS	KVSAV	TTLFV	PTVIN	TLGH	YVMP	ETFA	AP			
		250	260	270	280	290	300				

The complete length ORF104a nucleotide sequence <SEQ ID 403> is:

25	1	ATGGA	AAAC	AAGG	CCGCT	CCTAG	GGCTC	CGGTT	GGCAC	TTTGC	CGGCG
	51	GATG	ACGTGG	GGAAC	GCTGC	CGATT	CCCGT	GCGCG	CAGGTA	TTGAA	CTTTG
	101	TCGAT	CGCGC	GACG	CTGGTG	TGGGT	GCGTT	TTACG	CGTGC	GGCGG	CGGTA
	151	TTGTT	GTGTT	TGCT	GCGATT	GGGCG	GCGGG	CTGCG	CAGT	GGCGG	GATTT
	201	TTCTT	TGGTGC	TGCTT	CAGGCG	TGCTG	CTGCT	CGCGT	CGCGC	GCAT	TTCCG
	251	CAAAC	TTTGT	GCTG	ATTGCC	CAAGG	GCTGC	ATTAT	ATTTC	GCGCA	GCACG
30	301	ACGCA	GCTTT	TGTGG	CAGAT	TTCCG	CGCTT	ACGAT	GATTT	TTGCG	GTGTT
	351	GTTGG	GTGTT	AAAGC	ACCGA	TGACT	GCGCG	TCAGA	AAATC	GGCTT	GTGTT
	401	TGCTG	CTTGC	CGGTT	TGCTT	ATGTT	TTTTTA	ACGTA	AAAT	CGCGA	GCTTG
	451	TCGGG	TTTGG	GCGCG	CTATG	GAAGG	CGCTG	TTGCT	TGNTG	GCGCA	GCGAG
	501	TATGC	ATCAT	GCTGT	TATG	CCGTG	CGGCA	AAAGC	CTGCT	TCGGC	CGCAT
35	551	TCGGC	CGGCA	ACAGAT	CTG	CTGTT	GATTT	ATCGC	CGCAG	TCGGC	CGGCT
	601	TTCCG	CGCGT	TGCGG	AGCT	CGACG	AGATC	GGAAC	CTTGG	ACGCT	ACGCT
	651	GGCGT	GCGTT	TGTTT	TGCGT	ATTGC	CTGCT	GAAT	ACGTA	ATGCT	GTACG
	701	GCTCG	TTCG	CGAGG	CGTTG	AAACA	TGGG	AGGCT	TCCAA	AGTCA	CGCGG
	751	GTAAC	AACTC	TGCTC	CCCGT	GTTTA	CCGTA	ATATT	TTCTT	TGCTC	GGCGA
40	801	TTATG	TGATG	CCTGA	TACTT	TTGCC	CGCGC	GGAT	ATGAAC	GGTTG	GCGTT
	851	ATGCC	CGCGC	ACTGT	CTGCT	GTCGG	GCGTG	CGGTT	ACGCG	GCGGT	GCGG
	901	GACAG	GCTGT	TCAAA	CGCGC	CTAG					

This encodes a protein having amino acid sequence <SEQ ID 404>:

45	1	MENQR	PLLGF	ALALLA	AMTW	GTLPIA	VRQV	LKFVD	APTLV	VWVRF	TAAAV
	51	LFVLL	ALGGR	LPKRRD	FSWC	SFRLLL	LGVA	GISAN	FVLIA	QGLHY	ISPTT
	101	TOVLW	QISPF	TMIVVG	VLVF	KDRMTA	AQKI	GLVLL	LAGL	MFENDK	FGE
	151	SLG	AYAKG	VLLCA	AGSM	AWCN	VAQKLL	SAQFG	PQQIL	LLIYA	ASAAV
	201	FLP	FAELAH	IGSLD	GTALW	CFAYC	CLNTL	IGYGS	FGEAL	KHWEAS	KVSA
50	251	VTTLF	PVFTV	IFSLG	HVYM	PDTFA	PEPMN	GLGYA	NGALV	VGGAV	TAAV
	301	DRLF	KRR*								

ORF104a and ORF104-1 show 98.2% identity in 277 aa overlap:

		10	20	30	40	50	60
orf104a.pep	MENQR	PLLGF	ALALLA	AMTW	GTLPIA	VRQV	LKFVD
orf104-1	MENQR	PLLGF	ALALLA	AMTW	GTLPIA	VRQV	LKFVD
		10	20	30	40	50	60
orf104a.pep	LPKRRD	FSWCSF	RLLLG	VAGISAN	FVLIAQ	GLHYIS	PTTTQ
orf104-1	LPKRRD	FSWCSF	RLLLG	VAGISAN	FVLIAQ	GLHYIS	PTTTQ
		70	80	90	100	110	120
orf104a.pep	KDRMTA	AQKIGL	VLLLAG	LMFYFN	DKFGE	LSGLG	AYAKG
orf104-1	KDRMTA	AQKIGL	VLLLAG	LMFYFN	DKFGE	LSGLG	AYAKG
		130	140	150	160	170	180
orf104a.pep	KDRMTA	AQKIGL	VLLLAG	LMFYFN	DKFGE	LSGLG	AYAKG

ORF104 shows 93.9% identity over a 277aa overlap with a predicted ORF (ORF104.ng) from *N. gonorrhoeae*.

20	orf104.pep	MENQRPLLGLFALLAAMTWGTLPSYSRQVLKVFVDAPTLVWVRPTVAAAVLVFLALGGR	60
	orf104.ng	MENQRPLLGLFALLAAMTWGTLPSYSRQVLKVFVDAPTLVWVRPTVAAAVLVFLALGGR	60
25	orf104.pep	LPKRRDPSWCSEFLLLLGVAGISANFLVLAQGLHYISPTTTQVLQWISPTPTMIVGVGLVF	120
	orf104.ng	LPKRRDPSWHSFLLLLGVTCISANFLVLAQGLHYISPTTTQVLQWISPTPTMIVGVGLVF	120
	orf104.pep	KDRMTAAQKIGLVLLLAGLLMYFNQKFGELSGLGAYKGVLLCAAGSMAWVCNAVAOKLL	180
30	orf104.ng	KDRMTAAQKIGLVLLVGLMFFNDKFGELSGLGAYKGVLLCAAGSMAWVCYAVAOKLL	180
	orf104.pep	SAQFGPQIILLIYAASAAVFLPFAEPAHIGSMGDTLANVCIAVYCLNTNIICYGSFGEAL	240
	orf104.ng	SAQFGPQIILLIYAASAAVFLKPAEPAHIGSLDGTLANVCVYCLNTNIICYGSFGEAL	240
35	orf104.pep	KHWKASKVSATTTLLPFTVFTNTLLGHYVPEFTFAAP	277
	orf104.ng	KHWKASKVSATTTLLPFTVFTLSLGHYVPEFTFAAPDNNMNLGYGVGLVWVVGAVTAAVG	300

The complete length ORF104ng nucleotide sequence <SEQ ID 405> is predicted to encode a protein having amino acid sequence <SEO ID 406>:

45

1	MEHQRELLG	ALLAMMTW	GTLPFAVRQV	LKFVDAPTVL	WVRFTVAARV
51	LVFLALIGGR	LPKRROFSWH	KSRLLGLGVT	GISANFVLLA	QGLHYISPTT
101	TQWLAQKQ	PMIVGVGLV	DFRMTAAQKI	GLVLLVGLGL	MFNDKPGFEL
151	SLGAYAKGV	LCLCAQSHAW	VCYAVOCCIL	SAQFGPQIIL	LLTYAASAAV
201	FLIXAEPAHI	SGDGLTFLAW	CFPYCKNTLL	ICYGSGFQEL	KHWEASKVSA
251	VTTLPVLEPV	IFSLLLGHVYM	PDTFAAPDMN	GLGYVGLAVV	VGGAVTAAVG
301	DFPKR*				

Further work revealed the complete gonococcal nucleotide sequence <SEO ID 407>:

50	1	ATGCAAAACG	AGGACCCCGT	CCTAGTGGCT	CGCTTGGCAC	TTTTCGGCGC
	51	GATCAGCGTG	GGAGCGCTGC	CGATTCGCGT	CGCGCAGATT	TGAAGTTTGT
	101	TCGATCGGCG	GACGCTGTGT	TGGTGCGGTT	TTACCGTGGC	CGCGCGGATT
	151	TTCTTTGTTT	TGCTGCGATT	GGGCGCGGCT	TCGCGAGAC	CGCGGGATT
55	201	TTCTTGCGAT	TCGATTCAAGC	TGCTGCTGCT	CGCGCTTGAC	GGCAATTCGG
	251	CAAGCTGTGT	CGGCGGCGCT	CGGCGGCGCT	CGGCGGCGCT	CGGCGGCGCT
	301	ACCGAGCTGT	TGTGGCAGAT	TGCGCGCTTT	ACGATGATGT	TTTGTGCGCT
	351	GTGTTGTGTT	ATGAGCCCGA	tgcTTCGCCG	CGACAAATTC	GTTTGTGTTT
60	401	TGCTGCTGTG	CGGTTTGTtG	ATGCAAAAT	CGCGAGTTGT	CGCGAGCGAG
	451	TCGGGTTTGG	CGGCTGTATG	GAGGCGTATG	TTGCTGTGTG	CGCGAGCGAG
	501	TATGGCTTGG	GCTGTTTATG	CGCTGGGCGA	AAAGCTGTGT	TGGCGGCAAT
	551	TGGCGGCGCA	ACGAGATTCTG	CTGTGATTAT	ATGTCGCGTA	ttgcgcgcgctg
65	601	TTCCGCGGCT	CGGCGGCGCT	CGGCGGCGCT	CGGCGGCGCT	CGGCGGCGCT
	651	TCGGTGGGTT	TTGTTTGTGT	ATGTCGCTAT	GAATGCGGTA	ATGCGTATCG

-25-

701 GCTCGTTCGG CGAGGCGGTG AACATTGGG AGGCTTCCAA AGTCAGCGCG
 751 GTAACACCT TGCTCCCGCT GTTACCGTA ATATTTTCTT TGCTCGGCA
 801 TTATGTGATG CCGTACTT TTGCGCGCC GGATATGAAC GTTTGGGTT
 851 ATGTCGCGCG ACTGCTGTC GTGCGGGGTG CCGTTACGCC GCGGTGGGG
 901 GACGCGCGT TCAACGCGC CTAG

5

This corresponds to the amino acid sequence <SEQ ID 408; ORF104ng-1>:

1 MENQRPLLGF ALALLAAMTW GTLPIAVRQV LKFVDAPTLV WVRFTVAAAV
 51 LFLVLLALGGR LPKRRDFSWH SFRLLLLGVT GISANFVLA QGLHYISFTT
 101 TQVLWQISFF TMTVVGSHVF KDRMTAAQKI GLVLLVGLL MFDNKFGEI
 151 SGLGAYAKV LICAAGSMAW VCVAQAQKLL SAQFGPQIILL LLIYAASAV
 201 FLFPAEPAHI GSLDGTLAWV CFVYCCINTL IGYGSFGEAL KHWEASKVSA
 251 VTLLLPVFTV IFSLLGHYVM PDTFAAPDMN GLGVGALVW VGGAVTAAGV
 301 DRPFKRR*

10

ORF104ng-1 and ORF104-1 show 97.5% identity in 277 aa overlap:

15	orf104-1.pep	10	20	30	40	50	60
	MENQRPLLGFALALLAAMTWGTLPIAVRQVLKFVDAPTLVWVRFTVAAAVLFLVLLALGGR						
	orf104ng-1	10	20	30	40	50	60
	MENQRPLLGFALALLAAMTWGTLPIAVRQVLKFVDAPTLVWVRFTVAAAVLFLVLLALGGR						
20	orf104-1.pep	70	80	90	100	110	120
	LPKRRDFSWCSFRLLLLGVAGISANFVLAQGLHYISFTTQVLWQISFTTMTVVGVLV						
	orf104ng-1	70	80	90	100	110	120
	LPKRRDFSWHSFRLLLLGVAGISANFVLAQGLHYISFTTQVLWQISFTTMTVVGVLV						
25	orf104-1.pep	130	140	150	160	170	180
	KDRMTAAQKIGLVLLAGLLMFFNDKFGEISGLGAYAKVLLCAAGSMAWVCVAQAQKLL						
30	orf104ng-1	130	140	150	160	170	180
	KDRMTAAQKIGLVLLAGLLMFFNDKFGEISGLGAYAKVLLCAAGSMAWVCVAQAQKLL						
35	orf104-1.pep	190	200	210	220	230	240
	SAQFGPQIILLIYAASAVFLFPAEPAHIGSLDGTLAWVCFAYCCINTLIGYGSFGEAL						
	orf104ng-1	190	200	210	220	230	240
	SAQFGPQIILLIYAASAVFLFPAEPAHIGSLDGTLAWVCFAYCCINTLIGYGSFGEAL						
40	orf104-1.pep	250	260	270			
	KHWEASKVSAVTLLPVFTVIFSLGHYVMPDTFAAP						
	orf104ng-1	250	260	270	280	290	300
	KHWEASKVSAVTLLPVFTVIFSLGHYVMPDTFAAPDMNGLGVGALVWVGGAVTAAGV						

In addition, ORF104ng-1 shows significant homology with a hypothetical *H. influenzae* protein:

45 gi|573895 (U32769) hypothetical [Haemophilus influenzae] Length = 306
 Score = 237 bits (598), Expect = 8e-62
 Identities = 114/280 (40%), Positives = 168/280 (59%), Gaps = 8/280 (2%)

50 Query: 30 QRPKXXXXXXXXXXMTGTLPIAVRQVLKFVDAPTLVWXXXXXXXXXXXXXXXXXXXXX- 88
 Q+P M WG-LPIA+-QVL ++A +V+W P
 Sbjct: 3 QQLPLGTFALITAMAWGSLPIALKQVLSVMAQTIVVYRFIIAIVSLIALLAYKKQLFE 62

Query: 89 -KKRRDFSWHSFRLLLLGVAGISANFVLAQGLHYISFTTQVLWQISFTTMTVVGVLV 146
 K R ++W ++L+G V +++NF+L + L+YI P+ Q+ +S F M++ GVL+P
 55 Sbjct: 63 IMKVQYAW----IMLIGVIGLTSNFLFSSSINIEPSVAQIFHLSSFGMLICGVLF 118

Query: 147 KDRMTAAQKIXXXXXXXXXXFFNDKFGEISGLGAYAKVLLCAAGSMAWVCVAQAQKLL 206
 K+++ QKI +FFND+F +GL Y+ GV+L G++ WV Y +AQK+L
 60 Sbjct: 119 KEKLGHLKIGKIFLLLLIGLGFNDKFDFAAGLNQYSTGVILGVGALIVWAYGMAKIM 178

Query: 207 SAQFGPQIILLIYAASAVFLFPAEPAHIGSLDGTLAWCVFYCCINTLIGYGSFGEAL 266
 +P QQILL+Y A F+P A+ L LA +CF+YCCINTLIGYGS+ EAL
 Sbjct: 179 LRKFNSSQIILMMYLGCAIAFMFMADFSQVQEIT-PLALICFYCCINTLIGYGSYAEAL 237

65 Query: 267 KHWEASKVSAVTLLPVFTVIFSLGHYVMPDTFAAPDMN 306

W+ SKVS V TL+P+FT++FS + HY P FAAP++N
 Sbjct: 238 NRWDVSKVSVVITLIVLETLFLESHIAHYFSPADFAPELN 277

Based on this analysis, including the presence of a putative leader sequence and several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N. meningitidis* and *N. gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 48

The following partial DNA sequence was identified in *N. meningitidis* <SEQ ID 409>:

```

1   ATGGTAGCTC GTCGGGCTCA TAACCCGAAG GTCGTAGGTT CGAATCCTGT
51  .CCCGCAACC TAAITTCAAA CCCCTCGTT CAATGCCGAG GG.GTTTGT
101 T.TTGCCTGT TTCCTTTTC CCGCTTCCG CGGCTCCCGT TTTTTCGGG
151 ATTTTCTTTC CGCGCCGAAT ATCGGAACGG CAGACGCCGC TCTGTTTGGC
201 GTTGCAAAAT CAGSCAGTIT GGCTACAATC TTCCGATTG TCTTCAAGAA
251 AGCCAACCAT GCCGACCCTC CGTTTACCG AATCCGTGAG CAAACAAGAC
301 CTTGATGCTC TTGCTGAGTG GGCAAAAGCA AGTTACGGTG CAGAAAGTTG
351 CTGGAAACAG CTGTATCTGA ACGGTCysCC TTTGGGCAAC CTGTCGCCGG
401 AATGGGTGGA ACGCGTsmmA AAAGACTGGG AGGCAGGCTG CyCGGAGTCT
451 TCAGACGGCA TTTTTCGTAA TgCGACGGC TgGcGTGATA TGGcCGGAcy
501 cTTACAGCAC CTCGCCCTCG GTTGGCACTG TGCGGGGCTG TTGGACGst
20  551 GGCGCAACGA GTGTTTCGAC CTGACCAGCG GCGGCGGCAA CCCCTTGTTC
601 ACGCTCGaAc GCGCCSYTT mCGTCTKTC GGACTGCTCA GCGCGCCGCT
651 CCATCTCAAC GGTCTGACCG AATCGGACGG CCGATGGCAT TTCTGGATAG
701 CAGGCGCCAG TCCGCACAAA GAGTTCGATC CCACAAACT CACAAATACT
751 TCGCGGGGCG GTGTTTCCGG CGCGGAATG CCCTCTGAAG CGGTGTGTG
25  801 CGAAAGCAGC GAAGACGCGG GTTGTGATAA AACCTGTgTT CGCTCATCC
851 GCCCGGTATC GAGCTGCAC AGCTCGCGCT CCCTCAGCGG GGGGTACAC
901 AATGAAATCC TGATGTATT CGATGCCGTC CTGCCG...

```

This corresponds to the amino acid sequence <SEQ ID 410; ORF105>:

```

1   MVARRAHNFK VVGSNFXPAT XFQTPRFNAE XLVXLPVSCF LFPAASVFECR
51  IFLPAALSER QTVACLRLQI QAVWLQSSAL SSRKPTMTPT RFTEVSRSKQD
101 LDALFEWAKA SYGAESCWKT LYLNQXPLGN LSPFWVERVX KDWEAGCXES
151 SDGIFLNADG WPDMSGRQLH LALGWHCAGL LDGWRNECFD LTDGGGNFLF
201 TLERAXXRPX GLLSRVHLN GLTESDGRWH FWIGRRSPHK AVDPNKLNDT
25  251 XAGGVSGEM PSEAVCRESS EEAGLDKTLI PLIRPVSLQH SLRSVSRGVH
35  301 NEILYVFDV LF...

```

Further work revealed the complete nucleotide sequence <SEQ ID 411>:

```

1   ATGCCGACCG TCGGTTTTAC CGAATCCGTC AGCAACCAAG ACCTTGTATG
51  TCTGTTTCGAG TGGCGAAAAG CAAGTTACGG TGCAGAAATG TCGTGGAAAT
101 CGCTGTATCT GAACGGCTCTG CTTTGGGCA ACCTGTCCGC GGAATGGGTG
151 GAACGCGTCA AAAAAGACTG GGAGCGAGCG TGCTCGGATG CTTAGACAGG
201 CATTTTCTTG AATCGGGACG GCTGGCCTGA TATGGGCGCA GCCTTACAGC
251 ACCTCGCCCT CGGTTGGCAG TGTCGGGGCG TGTTGGACGG CTGGCGCAAC
301 GAGTGTTCG ACCTGACGCA GCGCGCGGCG AACCCCTTGT TACAGCTCGA
351 ACGCGCCGCT TTCCGTCTCT TCGGACTGCT CAGCCGCGCC GTCCATCTCA
45  401 ACGGCTGAC CGAATCGGAC GGCCGATGGC ATTTCTGGAT AGGCAGCGCG
451 AGTCCGCACA AAGCAGTCTG TCCCAACAAA CTCGACATA CTGCCGCCGG
501 CGGTGTTTCC GCGCGGAAAA TGCCGTCTGA AGCCGTGTGT CGCGAAAGCA
551 GCGAAGAAGC CGGTTTGGAT AAAACGCTGC TTTCCGCTCAT CCGCCCGGTA
601 TCGCAGCTGC ACAGCCTGCG CCGCTCGAG CGGGGTGTAC CAAATGAAAT
50  651 CTGTATCTGA TTGAGGCGCG TCTCGCCGGA AACCTTCTG CTTGAAATC
701 AGGATGGCGA AGTGGCGGGT TTTGAGAAAA TGGACATGGG CGGTCTGTG
751 GATGCAATGT TGTGGGAAA CATGATGAC ACACGCAAC TGGTACGCT
801 GGACGCGTTT TCCGTTTACG GTCTGATTTA TGCCGCCCAT CCGCTGTCCG
851 AGTGGCTGGA CGGCATACGT TTATAG

```

This corresponds to the amino acid sequence <SEQ ID 412; ORF105-1>:


```

101 ECFDLTDGGS NPLFALERAA FRPFGLLSRA VHNLGLVESD GRWHFWIGRR
151 SPHKAVDPDK LDNTAAGGVS SGELPSETVC RESSEEAAGL KTLPLIRPV
201 SQLHSRFPVS RGVHNEILYV FDAVLPEFEL PENQDGEVAG FEKMDIGGLL
251 AAMLSGNMMH DAQLVTLDFAF CRYGLIDAAH PLSEWLDGIR L*

```

5 ORF105a and ORF105-1 show 93.8% identity in 291 aa overlap:

```

10 orf105a.pep      10      20      30      40      50      60
      MPTVRFTEVSVKHDLDALEFWEAKASYGAESCWKTLTYNLGLPLGNLSPEWAERVKKDWEAG
10 orf105-1         10      20      30      40      50      60
      MPTVRFTEVSVKQDLDALEFWEAKASYGAESCWKTLTYNLGLPLGNLSPEWAERVKKDWEAG

15 orf105a.pep      70      80      90      100     110     120
      CSSESDGIFLNADGWPMGRRLLQHLARIWKEAGLLHGWRDECFLDTGGSSNPLFALERAA
15 orf105-1         70      80      90      100     110     120
      CSSESDGIFLNADGWPMGRRLLQHLALGWHCAGLLDGRWRNECFDLDTGGGNPLFTLERAA

20 orf105a.pep      130     140     150     160     170     180
      FRPFGLLSRAVHNLGLVESDGRWHFWIGRRSPHKAVDPDKLDNTAAGGVSSGELPSETVC
20 orf105-1         130     140     150     160     170     180
      FRPFGLLSRAVHNLGLTESDGRWHFWIGRRSPHKAVDPDKLDNTAAGGVSSGEMPEAVC

25 orf105a.pep      190     200     210     220     230     240
      RESSEEAAGLDKTLPLIRPVSQHLSLRPVSRGVHNEILYVFDVAVLPEFELPENQDGEVAG
25 orf105-1         190     200     210     220     230     240
      RESSEEAAGLDKTLPLIRPVSQHLSLRVSRRGVHNEILYVFDVAVLPEFELPENQDGEVAG

30 orf105a.pep      250     260     270     280     290
      FEKMDIGGLLAAMLSGNMMHDAQLVTLDFAFCRYGLIDAAHPLSEWLDGIRLX
30 orf105-1         250     260     270     280     290
      FEKMDIGGLDAMLSGNMMHDAQLVTLDFAFCRYGLIDAAHPLSEWLDGIRLX

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF105 shows 87.5% identity over a 312aa overlap with a predicted ORF (ORF105.ng) from *N.gonorrhoeae*:

```

40 orf105.pep      MVARRAHNPKVVGSNFPAFXQTFRFNAEXVLXLPVSCFLFPAASVFCRI FLPAALISER 60
      |||||
40 orf105ng        MVARRAHNPKVVGSNFPAFXQTFRFNAEGLVLF----FLFPAASVFCRI FLPAALISER 55

45 orf105.pep      QTAVCLRLQIQAVWLQSSALSSRKPTMTPTVRFTEVSVKQDLDALEFWEAKASYGAESCWK 120
45 orf105ng        QAQVCLRLQIQAVWLQSSALSSRKPTMTPTVRFTEVSVKQDLDALEFWEAKASYGAESCWK 115

      |||||
45 orf105.pep      LYLNKPLGNLSPEWVERVKKDWEAGCSESSDGIFLNADGWPMGRRLLQHLALGWHCAGL 180
45 orf105ng        LYLNKPLGNLSPEWVERVKKDWEAGCSESSNGIFLNADGWPMGRRLLQHLARTWNKAGL 175

50 orf105.pep      LDGWRNECFDLTDGGGNPLFTLERAXXRPXGLLSRAVHNLGLTESDGRWHFWIGRRSPHK 240
50 orf105ng        LHGWRNECFDLTDGGGNPLFTLERAAFRPFGLLIRAVHNLGLVESNGRWHFWIGRRSPHK 235

55 orf105.pep      AVDPNKLNTXAGGVSSGEMPEAVCRSESEEAAGLDKTLPLIRPVSQHLSLRVSRRGVH 300
55 orf105ng        AVDPKLDNTAGGVSSGEMPEAVCRSESEEAAGLDKTLPLIRPVSRLSLRVSRRGVH 295

60 orf105.pep      NEILYVFDVAFLP 312
60 orf105ng      NEILYVFDVAFLPETFLPENQDGEVAGFEKMDIGGLDAMLSGNMMHDAQLVTLDFAFIRYG 355

```

A complete length ORF105ng nucleotide sequence <SEQ ID 415> was predicted to encode a protein having amino acid sequence <SEQ ID 416>:

-257-

5
1 MVARRAHNFK VVGSNPAFAT KYQTPRFNAE GVLFLLFPAA SVFCRIFLPA
51 ATSERQAAYC LRLQICAVWL QSSALSRKP AMPTVRFTEA VSKQDLALF
101 ERAKASYGAE SCWKTYLYNR LPLGNLSPWE AERIKKDWEA GCSESSNGIF
151 LNADGWPMGM GRLQHLARTW NKAAGLHGW RNECDFLDGG GNPLFTLERA
201 AFRPFGLLIR AVHLNGLVES NGRWHFWIGR RSPHKAVDPG KLDNIAGGGV
251 SGSEMPSEAV CRESSEAGL DKTFLPLIRP VSRLHSIRPV SRGVHNEILY
301 VFDVLPETFL LPENQDGEVA GFEMDGGGL LDAMLKSNM HDAQLVTLDA
351 FYRYGLIDAA HFLSEWLDGI RL*

Further work revealed the complete nucleotide sequence <SEQ ID 417>:

10
1 ATGCGGACCG TCGCTTTTAC CGAATCGTC AGCAACAAAG ACCTTGATGC
51 CCGTGTGAGG CGGGCAAAAG CAAGTACGG TCGCGAAATG TGCTGGAAAT
101 CGCTGTATCT GAACCGCTCT CTTTGGGCA ATCTGTGCC GGAATGGGCT
151 GAGCGCATCA AAAAGACTG GAGGCGAGG TGCTCCGAGT CTTGAGACGG
201 CATTTTTCG AATGGGACG GCTGGCGGA TATGGGCGGA CGCTTGACG
15 ACCTGCGCCG CACATGGAAC AAGGCGGGG TGCTTCACGG ATGGCGCAAC
301 GAGTGTTCG ACCTGACCGA CGGCGGGGG AACCCCTTG TACGCTCGA
351 ACGCGCCGCT TTCCGTCGCT TCGGACTACT CAGCGCGGCC GTCCATCTCA
401 ACGTGTGGT CGAATCGAAC GGCAGATGGC ATTTTGTAGT AGCGAGCGCG
451 AGTCGCGACA AAGCAGTCGA TCOCGGCAAG CTCGACAATA TTGCGCGGG
20 CGGTGTTTCC GCGCGCGAAA TGCGCTCTGA AGCCGTGTGC CGCGAAAGCA
551 GCGAGAGAGC CGGTGTGGAT AAAAGCGCTG TTCCGCTCAT CCGCCAGATA
601 TCGCGCTGC ACAGCTCTGC CCGCTCGAG CGAGGTGTGC ACATGGAAT
651 CCGTGTATG TCGATGCGG TCGTCCGCGA AACCTTCG CCGTGAATC
701 AGGATGCGCA GGTGACGGGT TTTGAAAGA TGGACATGCG CGGCTATTG
25 GATGCGCATG TGTGCAAAAA CATGATGCAC GACCGCGAAC TGGTACGCT
801 GACGCGCTT TACCGTTACG GTCTGATTGA TGCCGCCCAT CCGCTGTCG
851 AGTGGCTGGA CGGCATACGT TTATAG

This corresponds to the amino acid sequence <SEQ ID 418; ORF105ng-1>:

30
1 MPTVRFTESV SKQDLALFE RAKASYGAE CWKTYLYNRL PLGNLSPWEA
51 ERIKKDWAG CSESSDGIFL NADGWPMGM RLQHLARTWN KAGLHGWNR
101 ECFDLTGGG NPLFTLERA RFPGLLSRA VHLNGLVES GRWHFWIGR
151 SPHKAVDPGK LDNIAGGGVS GSEMPSEAVC RESSEAGLD KTLFPLIRPV
201 SRLHSIRPV SRGVHNEILY VFDVLPETFL PENQDGEVAG FEKMDIGGL
251 DAMLSKMMH DAQLVTLDAF YRYGLIDAAH FLSEWLDGIR L*

35 ORG105ng-1 and ORF105-1 show 93.5% identity in 291 aa overlap:

		10	20	30	40	50	60
orf105-1.pep		MPTVRFTESVSKQDLALFEWAKASYGAESCWKTYLYNGLPLGNLSPWEVVERVKKDWAG					
orf105ng-1		MPTVRFTESVSKQDLALFERAKASYGAESCWKTYLYNRLPLGNLSPWEAERIKKDWAG					
		10	20	30	40	50	60
orf105-1.pep		70	80	90	100	110	120
orf105ng-1		CSESSDGI FLNADGWPMGMGRQLQHLALGWHCAGLLDGRWNECDFLDGGGNPLFTLERA					
		70	80	90	100	110	120
orf105-1.pep		130	140	150	160	170	180
orf105ng-1		FRPFGLLSRAVHLNGLTESDGRWHFWIGRRSPHKAVDPKLDNTAGGGVSGGEMPEAVC					
		130	140	150	160	170	180
orf105-1.pep		190	200	210	220	230	240
orf105ng-1		RESSEAGLDKTLPLIRPVSQLHSIRVSRGVHNEILYVFDVLPETFLPENQDGEVAG					
		190	200	210	220	230	240
orf105-1.pep		250	260	270	280	290	
orf105ng-1		FEKMDIGGLDAMLKSNMMHDAQLVTLDAFYRYGLIDAAHFLSEWLDGIRLX					
		250	260	270	280	290	

-25-

Furthermore, ORF105ng-1 shows homology with a yeast enzyme:

```

sp|P41888|TNR3_SCHPO THIAMIN PYROPHOSPHOKINASE (TPK) (THIAMIN KINASE)
>gi|1076925|t|S2350 thiamin pyrophosphokinase (EC 2.7.6.2) - fission yeast
(Schizosaccharomyces pombe) >gi|666111 (X84417) thiamin pyrophosphokinase
5 [Schizosaccharomyces pombe] >gi|2330852|gnl|PID|e334056 (Z98533) thiamin
pyrophosphokinase [Schizosaccharomyces pombe] Length = 569
Score = 105 bits (259), Expect = 4e-22
Identities = 64/192 (33%), Positives = 94/192 (49%), Gaps = 3/192 (1%)

10 Query: 268 NKAGLLHGWRNECFDLTDGGGNPLFTLERAARFPGLLSRAVHLNGLVESNGRW--HFWI 441
      N G+ WRNE ++ P+ +ER F EG LS VH ++ W+
Sbjct: 96 NTFGIADQWRNELYTVYGKSCKPVLAVERGGEWLFGLSTGVCHTMYIPATKEHFLRIWV 155

15 Query: 442 GRSPHKAVDPCKLDNIAGCGVSGGMPSEAVCRSESEAGLDKTLFFLPFVSRSLHSR 621
      RSP K F LDN GC++ G+ + +E SEEA LD + LI F + ++
Sbjct: 156 PRSPPTKQTWPNYLDNSVAGGIAGDTSVICTMIKEPSEANLDVSSMNLII-PCGTVSYIK 214

Query: 622 PVSRG-VHNEILYVFDVAVLPETFLPENQDGEVAGFEKMDIGLLDAMLKNNMMDAQLVT 798
      R + E+ YVFD ++ +P DGEVAGF ++ +L + K+ + LV
20 Sbjct: 215 NEKRHWIQPELQVYFDLPVDDLVIPIRINDGEVAGFSLPLNQVLELELKSFKPNCALVL 274

Query: 799 LDAFYRYGLIDAAHP 843
      LD R+G+I HP
Sbjct: 275 LDFLIRHGIIIPQHP 289

```

- 25 Based on this analysis, including the presence of a putative transmembrane domain in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 49

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID

```

30 419>:
      1 ATGAATAGAC CCAAGCAACC CTTCCTCCGT CCGGAAGTCG CCGTGGCCGC
      51 CCAAACACGAC CTGACGGGTA AACTGATCTG GACACGACCG TTGCTATTTT
      101 CCCTATGGAC GACATTGCA TCGATATCTG CGTTATTGAT TATCCTGTTT
      151 TTGATATTTG GTAACATATC GCGAAGACA ACAGTGGAGG GACAAATTTT
      201 ACCTGCATCG GCGCTAATCA GGCTCTATGC ACCGATACG rGkACAATTA
      251 CAGCGAATTT CGTGAAGAT GGGsAAAAAG TTAAGGCTGG CGACAAGCTA
      301 TTTCGGCTTT CGACCTCAAG TTTCGGCGCA GGAGGTAGCG TGCAGCAGCA
      351 GTTGAAACCG GAGGCAGTTT TGAAGAAACG GTTGGCAGAA CAGGAACCTG
      401 CTCGTCTGAA GGTGATACAC GGAAGTGAAG CGCGAGuCT TAAGCAACT
      451 CTCGAACGTT TGGAAACCA GGAACCTCAT ATTTGCAAC AGATAGACCG
      501 TCAGAAAGAG CGCATTAGAC TTGCGGAAGA AATCTTGACG AATATCGTT
      551 TCCTATCCGC .CAATGA

```

This corresponds to the amino acid sequence <SEQ ID 420; ORF107>:

```

45      1 MNRPKQPFPR PEVAVARQTS LTGKVLTRP LSFSLWTTFA SISALLIILF
      51 LIFGNVTRKT TVEGQILPAS GVIRVYAPDT XTITAKFYED GKKVKAGDKL
      101 FALSTRSRGA GGSVQQQLKT EAVLKTKLAE QELGRKLKH GNETRSLKAT
      151 VERLENQELH ISQQIDGQKR RIRLAEMLQ KYRFLSXQ*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

- 50 ORF107 shows 97.8% identity over a 186aa overlap with an ORF (ORF107a) from strain A of *N.meningitidis*:

10 20 30 40 50 60

-259-

orf107.pep	MNRPKQPFRRPEVAVARQTSLTGKVLTRPLSFSLWTTFASISALLIILFLIFGNYTRKT	
orf107a	MNRPKQPFRRPEVAVARQTSLTGKVLTRPLSFSLWTTFASISALLIILFLIFGNYTRKT	
	10 20 30 40 50 60	
orf107.pep	70 80 90 100 110 120	
orf107a	70 80 90 100 110 120	
	130 140 150 160 170 180	
orf107.pep	EAVLKKTAEQELGRKLHNGNETRSLKATVERLENQELHISQQIDGKRRIRLAEEMLQ	
orf107a	EAVLKKTAEQELGRKLHNGNETRSLKATVERLENQELHISQQIDGKRRIRLAEEMLQ	
	130 140 150 160 170 180	
orf107.pep	189 KYRFLSXQX	
orf107a	KYRFLSANDAVFKQEMMNVKAELEQKAKLDAYRREEVGLLQEIQTQNLTLXSLPQAA	
	190 200 210 220 230	

The complete length ORF107a nucleotide sequence <SEQ ID 421> is:

25	1	ATGAATGAC	CCAAGCAAC	NTTCTCCGT	CCGGAAGTCG	CGTGTGCCG
	51	CCAAACGAGC	CTGACGGSTA	AAGTGATTCT	GACACGACCG	TTGTCAATTT
	101	CCCTATGGAC	GACATTTGCA	TGCAATCTGT	CGTTATTTAT	TATCTGTTT
	151	TTGATATTGT	GTAACATATC	GCGAAGACA	ACAGTGGAG	GACAAATTT
	201	ACCTGCATCG	GGCGTAATCA	GGGTGTATCG	ACGGATACG	GGGACAATTA
	251	CNGCGAAATT	CNTGGAAGAT	GGGAAAAGG	TTAAGGCTGG	CGACAACTTA
30	301	TTTGCGCTTT	CGACCTCACG	TTTCGCGCGA	GGAGATAGCG	TGCACGACGA
	351	GTGGAAGACG	GAGGCGATTT	TGAAGAAAAC	GTTCGCGAGAA	CAGGAACCTGG
	401	TCGTCTGAA	GCTGATACAC	GGGAATGAAA	CGCGACGCTT	TAAAGCAACT
	451	TCGGAACGTT	TGGAAAACCA	GGAATCTCAT	ATTTCGACAG	AGATAGACGG
	501	TCGGAAGAGG	CGCAATTAGAC	TTCCGGAAGA	AATGTTGCGC	AAATATCGTT
35	551	TCCATTCGCG	CAATGATGCA	GTCGCAAAAC	AGAAATGAT	GAATGTCAG
	601	GCAGAGCTTT	TAGAGCAGAA	AGCCAAACTT	GATGCTACC	GCGAGAGA
	651	AGTGGGCTG	CTTCACGAAA	TCCGCAACGA	GAATCTGACA	TTGNNAGCC
	701	TCCCCCAAGC	GGCATGA			

This encodes a protein having amino acid sequence <SEQ ID 422>:

40	1	MNRPKQPFRR	PEVAVARQTS	LTGKVLTRP	LSFSLWTTFA	SISALLIILF
	51	LIFGNYTRKT	TVEGQILPAS	GVIRVYAPDT	GTITAKFVED	GERVKVAGDKL
	101	FALSTSRFGA	GDSVQQQLKT	EAVLKKTAE	QELGRKLH	NETRSLKAT
	151	VERLENQELH	ISQIDGQKR	RIRLAEEMLQ	KYRFLSANDA	VFKQEMMNVK
	201	AELEQKAKL	DAYRREEVGL	LQEIQTQNL	LXSLPQAA*	

Homology with a predicted ORF from *N.gonorrhoeae*

ORF107 shows 95.7% identity over a 188aa overlap with a predicted ORF (ORF107.ng) from *N.*

gonorrhoeae:

50	orf107.pep	MNRPKQPFRRPEVAVARQTSLTGKVLTRPLSFSLWTTFASISALLIILFLIFGNYTRKT	60
	orf107.ng	MNRPKQPFRRPEVAVAIARQTSLTGKVLTRPLSFSLWTTFASISALLIILFLIFGNYTRKT	60
	orf107.pep	TVEGQILPASGVIRVYAPDTXTITAKFVEDGKVKAGDKLFALSTSRFGAGGSVQQQLKT	120
55	orf107.ng	TMEGQILPASGVIRVYAPDTGTTITAKFVEDGKVKAGDKLFALSTSRFGAGGSVQQQLKT	120
	orf107.pep	EAVLKKTAEQELGRKLHNGNETRSLKATVERLENQELHISQQIDGKRRIRLAEEMLQ	180
	orf107.ng	EAVLKKTAEQELGRKLHNGNETRSLKATVERLENQELHISQQIDGKRRIRLAEEMLR	180
60	orf107.pep	KYRFLSXQ 188	
	orf107.ng	KYRFLSAQ 188	

The complete length ORF107ng nucleotide sequence <SEQ ID 423> is predicted to encode a protein having amino acid sequence <SEQ ID 424>:

```

1  MNRPKQFFFR PEVAIAQRQTS LTGKVILTRP LSFSLWTTFA SISALLIILF
51  LIFGNTRYRKT TMEGQILPAS GVRVYAPDT GTITAKFVED GERVKAGDKL
5  101  FALSTSRFGA GGSVQQLKLT EAVLKKTLEL QELGRILKLIH ENETRSLKAT
151  VERLENQKLH ISQQIDGQKR RIRLAEEMLR KYRFLSAQ*

```

Based on the presence of a putative ransmembrane domain in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

10 Example 50

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID 425>:

```

1  ATGCTGAATA CTTTTTTTGC GGTATTGGGC GGCTGCCCTGC TGCT. TTGCC
15 51  GTGGCGGCAAA TCCGTAATAA CGCGSGTACA GCGGCAAAAC GCGGTACAAA
101  GCGCGCGCGAA ACCGGTTTTT AAAGTCATAT ATATCGACAA TAGCGCGATT
151  GCGCGTTTTGG ATTTGGGACA AAGCAGCGAA GCGCAAAACA ACAGCGCGAA
201  AAAACAARTC AGTTATCCGA TTAAGGCTT GCGGGAACAA AATGTTATCC
251  GACTGATCGG CAAGCATCCC GCGCATTTGG AAGCGGTGAG CGGCAAAATG
301  ATGGAAACCG ATGATAAGGA CAGTCCGCGA GGTGGGCGAG AAAACGGCGT
20 351  GTGCCATACC TTGTTTGCCA AACTGGTGGG CAATATCGCC GAAGACGGCG
401  GCAAACTGAC GATTACCTTA GTTTCGCAAT GCGCGCTGCA ACCCTATCAG
451  GCAGGCAAAA GCGGCTATGC CGCGTGCAG AAGGAGCGCT ATGTGCTGGA
501  AATGACAGC GAAGGGGCGT TTTATTTCCG CCGCGGCAT TATTGA

```

This corresponds to the amino acid sequence <SEQ ID 426; ORF108>:

```

25 1  MLNTFFAVLG GCLLXLPGK SVNTAVQPN AVQSAPKPVF KVIYIDNTAI
51  AGLDLGQSSE GKTNDGKKQI SYPIKGLPEQ NVIRLIGKHE GDLEAVSGKC
101  METDDKDSFA GWAENGVCHT LFALKVGNIA EDGKILTDYL VSHAALQPYQ
151  AGKSGYAAVQ NGRYVLEIDS EGAFYFRRRH Y*

```

Further work revealed the following DNA sequence <SEQ ID 427>:

```

30 1  ATGCTGAAAA CATCTTTTGC GGTATTGGGC GGCTGCCCTGC TGCTTGCGCG
51  CTGCGGCAAAA TCCGTAATAA CGCGGGAACA GCGGCAAAAC GCGGTACAAA
101  GCGCGCGCGAA ACCGGTTTTT AAAGTCAAAT ATATCGACAA TAGCGCGATT
151  GCGCGTTTTGG ATTTGGGACA AAGCAGCGAA GCGCAAAACA ACAGCGCGAA
201  AAAACAARTC AGTTATCCGA TTAAGGCTT GCGGGAACAA AATGTTATCC
35 251  GACTGATCGG CAAGCATCCC GCGCATTTGG AAGCGGTGAG CGGCAAAATG
301  ATGGAAACCG ATGATAAGGA CAGTCCGCGA GGTGGGCGAG AAAACGGCGT
351  GTGCCATACC TTGTTTGCCA AACTGGTGGG CAATATCGCC GAAGACGGCG
401  GCAAACTGAC GATTACCTTA GTTTCGCAAT GCGCGCTGCA ACCCTATCAG
451  GCAGGCAAAA GCGGCTATGC CGCGTGCAG AAGGAGCGCT ATGTGCTGGA
40 501  AATGACAGC GAAGGGGCGT TTTATTTCCG CCGCGGCAT TATTGA

```

This corresponds to the amino acid sequence <SEQ ID 428; ORF108-1>:

```

45 1  MLKTSFAVLG GCLLLAACGK SENTAEQPN AVQSAPKPVF KVIYIDNTAI
51  AGLDLGQSSE GKTNDGKKQI SYPIKGLPEQ NVIRLIGKHE GDLEAVSGKC
101  METDDKDSFA GWAENGVCHT LFALKVGNIA EDGKILTDYL VSHAALQPYQ
151  AGKSGYAAVQ NGRYVLEIDS EGAFYFRRRH Y*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.gonorrhoeae*

ORF108 shows 88.4% identity over a 181aa overlap with a predicted ORF (ORF108.ng) from *N.*

gonorrhoeae:

5	orf108.pep	MLNTFFAVLGGCLLXLPCGKSNTAVQPONAVQSAKPVFKVIYIDNTAIAGLDLGQSSSE	60
	orf108.ng	MLKIPFAVLGGCLLLAARCGKSENTAEQPNAAQSAKPVFKVIYIDNTAIAGLDLGQSSSE	60
10	orf108.pep	GKTDNGKKQISYPKGLPEQNVIRLIGKHPGDLAVSGKCMETDDKDSFAGWAENGVCHT	120
	orf108.ng	GKTDNGKKQISYPKGLPEQNAVLTKGHPNDLEAVVGKCMETDGKDAFSGWAENGVCHT	120
15	orf108.pep	LFAKLVGNIAEDGGKLTDLVLSHAALQPYQAGKSGYAAVQNGRYVLEIDSEGAFFRRHHY	181
	orf108.ng	LFAKLVGNIAEDGGKLTDLVLSHAALQPYQAGKSGYAAVQNGRYVLEIDSEGAFFRRHHY	181

ORF108-1 shows 92.3% identity with ORF108.ng over the same 181 aa overlap:

20	orf108-1.pep	MLKTSFAVLGGCLLLAARCGKSENTAEQPNAAQSAKPVFKVIYIDNTAIAGLDLGQSSSE	60
	orf108.ng-1	MLKIPFAVLGGCLLLAARCGKSENTAEQPNAAQSAKPVFKVIYIDNTAIAGLDLGQSSSE	60
25	orf108-1.pep	GKTDNGKKQISYPKGLPEQNVIRLIGKHPGDLAVSGKCMETDDKDSFAGWAENGVCHT	120
	orf108.ng-1	GKTDNGKKQISYPKGLPEQNAVLTKGHPNDLEAVVGKCMETDGKDAFSGWAENGVCHT	120
30	orf108-1.pep	LFAKLVGNIAEDGGKLTDLVLSHAALQPYQAGKSGYAAVQNGRYVLEIDSEGAFFRRHHY	181
	orf108.ng-1	LFAKLVGNIAEDGGKLTDLVLSHAALQPYQAGKSGYAAVQNGRYVLEIDSEGAFFRRHHY	181

The complete length ORF108.ng nucleotide sequence <SEQ ID 429> is:

30	1	ATGCTGAAAA	tacctTTTGC	CGTGTtgggc	ggCtgccctgc	TGCTTGCCGC
	51	CTGCGCGCAA	TCGGAATA	cggcggAAAT	GCGCAAAAT	gCGGCAAAA
35	101	GCGCGCGCAA	ACCGGTTTT	AAAGTCAAT	ACATCGACAA	TACGCGGATT
	151	GCGCGTTTTG	CTTTGGGACA	AAGTAGCGAA	GGCAAAACCA	acgaacGCAA
40	201	AAACCAAAAT	AGTTATccga	TTAAAGGCTT	GCGCGAACAA	Aacgcgctcc
	251	gGCTGACCGG	AAGCATCCO	AACGACTTGG	Aagcgcgtcg	CGGCAAAATG
45	301	ATGGAACCG	ACGGAAAGGA	CGCGCTTCG	GGCTGGGCGG	AAACCGCGGT
	351	GTGCCATACC	TTGTTTGCCA	AACGTGTGGG	CAATATCGCG	GAAGACGGCG
50	401	GCAAACTGAC	TGATTACTCG	ATTTCGCGAT	CGCGCTGCA	ACCCATACAG
	451	GACGGCAAAA	GCGGCTATGC	GCGCGTCGAG	AACGACGCT	ATGTGCTGGA
	501	AATCGACAGC	GagggGGCGT	TTTATttccg	cgcgcgcat	tattgA

40 This encodes a protein having amino acid sequence <SEQ ID 430>:

45	1	MLKIPFAVLG	GCLLLAARCGK	SENTAEQPN	AAQSAKPVF	KVKYIDNTAI
	51	AGLALGQSSSE	GKTDNGKKQI	SYPIKGLPEQ	NAVLTKGHP	NDLEAVVGKC
50	101	METDGKDAFS	GWAENGVCHT	LFAKLVGNIA	EDGKLTDLV	LSHSAALQPYQ
	151	AGKSGTAAVQ	NGRYVLEIDS	EGAFYFRHHY	Y*	

Based on this analysis, including the presence of a predicted prokaryotic membrane lipoprotein lipid attachment site (underlined) and a putative ATP/GTP-binding site motif A (P-loop, double-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

50 Example 51

The following DNA sequence was identified in *N.meningitidis* <SEQ ID 431>:

-262-

```

1  ATGGAAGATT  TATATATAAT  ACTCGCTTGT  GGTTTGGTTG  CGATGATTGC
51  CGGATTATC  GATGcgatTg  cggcggggg  TGGTTTGATT  ACGTGCCCGG
101  CACTCTTGTT  GGCAGGTATT  CTTCCCGTGT  CGGCAATTGC  CACCACACAG
151  CTGCAAGCAG  CCGCTGCTAC  GTTTTCAGCT  ACGGTTTCTT  TTGCACGCAG
201  AGGTTTGATT  GATTGGAAGA  AAGGTCCTCC  GATTGCGCGA  GCATCGTTTG
251  TAGGCGGCGT  GgcCGGTGCA  TTATCGGTCA  GCTTGGTTTC  CAAAGATATT
301  CTGCTgCGG  TCGTGCCGGT  TTTGTTGATA  TTTGTCGCAC  TGTATTTTGT
351  GTTTTCGCC  AAGCTCGACG  CGAGTAAGGA  AGGCAAGGCC  AGAATGTCTT
401  TTTTTCGTGT  CCGGCTGACG  GTCGC  ACGC  CTTTGGGTTT  TTACGACGG
451  TCGTTCGGA  CCGGCTGCG  GCTCGTTT  TCGATAGCC  TTTATGTTT
501  TGTCTGGCTG  CAAgCTGTTG  AACCGATGT  CTACACCAA  ATTGGCGAAC
551  GTTGCTGCA  ATCTTGGTTC  GCTATCGTA  TTCTGCTGC  ACGGTTTCGAT
601  TATTTTCCG  ATTGCGGCAG  CGATGCGGT  CGTGCGTGT  GTCGGTGCGA
651  ATTTAgGTGC  GAGATTTGCC  GTaCgctTCG  GTTCGAAGCT  GATTAA

```

15 This corresponds to the amino acid sequence <SEQ ID 432; ORF109>:

```

1  MEDLYIILAL  GLVAMIAGFI  DAIAGGGGLI  TLPALLLAGI  PPVSAIATNK
51  LQAAAATFSA  TVSFARKGLI  DWKKGLPIAA  ASFVGGVAGA  LSVSLVSKDI
101  LLAVVPVLLI  FVALYFVFSF  KLDGSKGKA  RMSFFLFGLT  VXTAFGLLR
151  CVRTGCRLVF  SDLCYCFARL  QAVERVVLHQ  IGERCLQSWF  AIGTAAFRFD
201  YFPDGNDDG  RCVCRCFRC  EICRTLRFEA  D*

```

Further work revealed the following DNA sequence <SEQ ID 433>:

```

1  ATGGAAGATT  TATATATAAT  ACTCGCTTGT  GGTTTGGTTG  CGATGATTGC
51  CGGATTATC  GATGCGATTG  CGGCGGGGG  TGGTTTGATT  ACGTGCCCGG
101  CACTCTTGTT  GGCAGGTATT  CTTCCCGTGT  CGGCAATTGC  CACCACACAG
151  CTGCAAGCAG  CCGCTGCTAC  GTTTTCAGCT  ACGGTTTCTT  TTGCACGCAG
201  AGGTTTGATT  GATTGGAAGA  AAGGTCCTCC  GATTGCGCGA  GCATCGTTTG
251  TAGGCGGCGT  GGCCTGCTCA  TTATCGGTCA  GCTTGGTTTC  CAAAGATATT
301  CTGCTGCGG  TCGTGCCGGT  TTTGTTGATA  TTTGTCGCAC  TGTATTTTGT
351  GTTTTCGCC  AAGCTCGACG  CGAGTAAGGA  AGGCAAGGCC  AGAATGTCTT
401  TTTTTCGTGT  CCGGCTGACG  GTCGCACCGC  TTTTGGGTTT  TTACGACGGT
451  GTGTTGCGAC  CCGGCTGCG  CTCGTTTTT  CTGATTGCGT  TTTATGTTT
501  GTCGCGCTG  AAGCTGTTGA  ACGCGATGC  TTACACCAA  TTGGCGAAGC
551  TTGCTGCAA  TCTTGGTTCG  CTATCGGTAT  TCCTGCTGCA  CGGTTTCGAT
601  ATTTTCCCG  TTGCGGCAAC  GATGCGGCT  GGTGGGTTTG  TCGGTGCGAA
651  TTTAGGTGCG  AGATTGCGG  TCGGCTCGG  TCGGAAGCTG  ATTAAGCCCG
701  TGTGATTGT  CATCAGATT  TCGATGGCTG  TGAATGTGT  GATAGACGAG
751  AGAAATCCGC  TGTATCAGAT  GATTGTTTCG  ATGTTTTAA

```

This corresponds to the amino acid sequence <SEQ ID 434; ORF109-1>:

```

1  MEDLYIILAL  GLVAMIAGFI  DAIAGGGGLI  TLPALLLAGI  PPVSAIATNK
51  LQAAAATFSA  TVSFARKGLI  DWKKGLPIAA  ASFVGGVAGA  LSVSLVSKDI
101  LLAVVPVLLI  FVALYFVFSF  KLDGSKGKA  RMSFFLFGLT  VAPLLGFDG
151  VFGPGVGSFF  LIAFTVLLGC  KLLNAMSYTE  LANVACNLGS  LSVFLHGSII
201  IFPIAATMAV  GAFVGNLGA  RFVRFSGSKL  TKPLLIVISI  SMAVKLLIDE
251  RNPLYQMIVS  MF*

```

45 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF109 shows 95.9% identity over a 147aa overlap with an ORF (ORF109a) from strain A of *N.*

meningitidis:

```

50  orf109.pep  MEDLYIILALGLVAMIAGFIDAIAGGGGLITLPALLLAGIPPVSAIATNKLQAAAATFSA
      orf109a  MEDLYIILALGLVAMIAGFIDAIAGGGGLITLPALLLAGIPPVSAIATNKLQAAAATFSA
      10      20      30      40      50      60
55  orf109.pep  TVSFARKGLIDWKKGLPIAAASFVGGVAGALSLSVSLVSKDILLAVVPVLLIFVALYFVFSF
      orf109a  TVSFARKGLIDWKKGLPIAAASFVGGVAGALSLSVSLVSKDILLAVVPVLLIFVALYFVFSF
      70      80      90      100     110     120

```

-263-

		130	140	150	160	170	180
orf109.pep		KLDGSKGKARMSFFLGLTVXTAFGLRRCVTRGLVFSDCLYCFARLQAVRDLVHQ					
5 orf109a		KLDGSKGKARMSFFLGLTVAPLLGFYDVGVPGVGSFFLIATFIVLLGCKLLNAMYTK					
		130	140	150	160	170	180

The complete length ORF109a nucleotide sequence <SEQ ID 435> is:

1	ATGGAAGATT	TATACATAAT	ACTCGCTTTG	GCTTGCTTG	CGATGATTGC
5	CGGATTATTC	GATCGGATGG	CGGCTGGGGC	TGGTTGATTT	ACGCTGCTGC
10	CACCTCTTGT	GGCAGCTATT	CCTCCGCTGT	CGCARTTGC	CACCAACAAG
15	CTGCAAGCAG	CGCGTCTGAC	GTTTTCGGCT	ACGTTTCTCT	TTGCACGCAA
20	AGGTTTGATT	GATTGGAAGA	AAGGTCTCCC	GATTGCGGCA	GCATCGTTTG
25	CAGGCGGCGT	GTCGCTGCA	TTATCGGTCA	GCTTGGTTTC	CAGAATATTT
30	CTGCTGGCGG	TGCTGCGGTT	TTTGTGATA	TTTGTGCGGC	TGTATTTTGT
35	GTTTTCGCC	AAGCTCGACG	GCAGTAAGGA	AGGCAAGCC	AGAATGCTCT
40	TTTTCTGT	CGGCTGACG	GTTGCACCA	TTTTGGTTT	TTACGACGGT
45	GTGTTGCGAC	CGGCTGCGG	CTCGTTTTT	CTGATTGCCT	TTATTGTTTT
50	GCTCGGCTGC	AAGCTGTTGA	ACGCGATGTC	TTACACCAA	TTGGCGAAGC
55	TTGCTCGCAA	CTCTGGTTG	CTATCGGTAT	TCCTGTGCA	CGGTCGATT
60	ATTTCCTCGA	TTGCGGCAAC	GATGCGGCTC	GCTGCGTTG	TGCGTGCAGAA
65	TTTAGGTCG	AGATTGCGG	TCCGCTTCGG	TTCAAGCTG	ATTAGCGCGC
70	TGCTGATTG	CTACGCAATT	TGCGATGCGT	TGAATGCT	GATAGACGAG
75	AGAAATCCG	TGATCAGAT	GATTGTTTC	ATGTTTTTA	

This encodes a protein having amino acid sequence <SEQ ID 436>:

1	MEDLYIILAL	GLVAMIAGFI	DAIAGGGGLI	TLPALLLAGI	PPVSAIATNK
5	LQAAAAATFSA	TVSFARKGLI	DWKKGLPIAA	ASFAGGVVGA	LSVSLVSKDI
10	LLAVVFPVLLI	FVALYFVFSP	KLDGSKGKA	RMSFFLGLT	VAPLLGFYD
15	VFGPGVGSFF	LFIATFVLLG	KLLNAMYTK	LANVACNLGS	LSVFLHGS
20	IFPIAATMAV	GAFGVANLGA	RFAVRFGSKL	IKPLLVISI	SMAVKLLIDI
25	RNFLYQIVS	MF*			

ORF109a and ORF109-1 show 99.2% identity in 262 aa overlap:

		10	20	30	40	50	60
orf109a.pep		MEDLYIILALGLVAMIAGFIDAIAGGGGLITLPALLLAGI	PPVSAIATNKLQAAAAATFSA				
35 orf109-1		MEDLYIILALGLVAMIAGFIDAIAGGGGLITLPALLLAGI	PPVSAIATNKLQAAAAATFSA				
		10	20	30	40	50	60
		70	80	90	100	110	120
orf109a.pep		TVSFARKGLIDWKKGLPIAAASFAGGVVAGLSVSLVSKDILLAVVFPVLLIFVALYFVFSP					
40 orf109-1		TVSFARKGLIDWKKGLPIAAASFAGGVVAGLSVSLVSKDILLAVVFPVLLIFVALYFVFSP					
		70	80	90	100	110	120
		130	140	150	160	170	180
orf109a.pep		KLDGSKGKARMSFFLGLTVAPLLGFYDVGVPGVGSFFLIATFIVLLGCKLLNAMYTK					
45 orf109-1		KLDGSKGKARMSFFLGLTVAPLLGFYDVGVPGVGSFFLIATFIVLLGCKLLNAMYTK					
		130	140	150	160	170	180
		190	200	210	220	230	240
orf109a.pep		LANVACNLGSLSVFLHGSIIIFPIAATMAVGAFVGNLGA	RFAVRFGSKLIKPLLVISI				
50 orf109-1		LANVACNLGSLSVFLHGSIIIFPIAATMAVGAFVGNLGA	RFAVRFGSKLIKPLLVISI				
		190	200	210	220	230	240
		250	260				
orf109a.pep		SMAVKLLIDERNPLYQIVSMFX					
60 orf109-1		SMAVKLLIDERNPLYQIVSMFX					
		250	260				

Homology with a predicted ORF from *N.gonorrhoeae*

ORF109 shows 98.3% identity over a 231aa overlap with a predicted ORF (ORF109.ng) from *N. gonorrhoeae*:

5	orf109.pep	MEDLYIILALGLVAMIAGFIDAAGGGGLITLPALLLAGIPPVSAIATNKLQAAATFSA	60
	orf109.ng		60
10	orf109.pep	MEDLYIILALGLVAMIAGFIDAAGGGGLITLPALLLAGIPPVSAIATNKLQAAATFSA	60
	orf109.ng		60
15	orf109.pep	TVSFARKGLIDWKKGLPIAAASFAGGVVAGALSVSLVSKDILLAVVPVLLIFVALYFVFS	120
	orf109.ng		120
20	orf109.pep	TVSFARKGLIDWKKGLPIAAASFAGGVVAGALSVSLVSKDILLAVVPVLLIFVALYFVFS	120
	orf109.ng		120
25	orf109.pep	KLDGSKEGKARMSFFLPGTLVXTAFGFLRRCVTRGCRVFSDCLYCFARLQAVRDVLHQ	180
	orf109.ng		180
30	orf109.pep	KLDGSKEGKARMSFFLPGTLVXTAFGFLRRCVTRGCRVFSDCLYCFARLQAVRDVLHQ	180
	orf109.ng		180
35	orf109.pep	IGERCLOSWFAIGIPAARFDYFFDCGNDGRCVCRCEFRCEICRTRLEAD	231
	orf109.ng		231

An ORF109ng nucleotide sequence <SEQ ID 437> was predicted to encode a protein having amino

acid sequence <SEQ ID 438>:

1	MEDLYIILALGLVAMIAGFIDAAGGGGLITLPALLLAGIPPVSAIATNKLQAAATFSA
51	TVSFARKGLIDWKKGLPIAAASFAGGVVAGALSVSLVSKDILLAVVPVLLIFVALYFVFS
101	KLDGSKEGKARMSFFLPGTLVXTAFGFLRRCVTRGCRVFSDCLYCFARLQAVRDVLHQ
151	IGERCLOSWFAIGIPAARFDYFFDCGNDGRCVCRCEFRCEICRTRLEAD
201	YFPDCGNDGGRCVCRCEFRCEICRTRLEAD

Further work revealed the following gonococcal DNA sequence <SEQ ID 439>:

1	ATGGAAGATT	TATACATAAT	ACTGCGTTTG	GGTTTGGTTG	CGATGATCGC
51	CGGATTTATC	GATGCGATTG	CGGCGGGGGG	TGGTTTGATT	ACGCTGCCTG
101	CACCTCTGTT	GCCAGGTATT	CCTCCCGTGT	CGGCAATTGC	CACCAACAAG
151	CTGCAAGCAG	CGCGTGCTAC	GTTTTCGGCT	ACGGTTTCTT	TGCGACGCAA
201	AGGTTTGATT	GATTGGAAGA	AAGGTCCTCC	GATGCGCGCA	GCATCGTTTG
251	CAGCGCGCGT	GGTCCGTGCA	TTATCGGTCA	GCTTGCTTTC	CAAGATATT
301	TGCTGCGCGG	TGCTGCGGT	TTTGCTGATA	TTTCTGCGCT	TGATTTTGT
351	GTTTTCGCCC	AGGCTCGACG	GCAGTAAGGA	AGGCAAGCC	AGAATGTCTT
401	TTTTTCTATT	CGGGCTGACG	GTGACCCGC	TTTTGGTTT	TTACGACGGT
451	GTGTTCCGAC	CGGGTGTGCG	CTGCTTTTT	CTGATTGCGT	TTATTGTTTT
501	GCTCGGCTGC	AGGCTGTGGA	ACGCGATGTC	TTACACCAAA	TTGCGCAACG
551	TTGCTTCCAA	TCTTGGTTTC	CTATCGGAT	TCCTGCTGCA	CGGTTGCGAT
601	ATTTTCCCGA	TTGTGGCAAC	GATGGCGGTC	GGTGGTTTTC	TCGGTGCAGAA
651	TTTAGTTCGG	AGATTTCGCG	TCCGCTTCGG	TTGAAAGCTG	ATTAAGCCCG
701	TGCTGATTGT	CATCAGCATT	TGCGATGCTG	TGAATTTGTT	GATAGACGAG
751	AGAAATCCGC	TGTATCAGAT	GATTGTTTCG	ATGTTTTTAA	

This corresponds to the amino acid sequence <SEQ ID 440; ORF109ng-1>:

1	MEDLYIILALGLVAMIAGFIDAAGGGGLITLPALLLAGIPPVSAIATNKLQAAATFSA
51	TVSFARKGLIDWKKGLPIAAASFAGGVVAGALSVSLVSKDILLAVVPVLLIFVALYFVFS
101	KLDGSKEGKARMSFFLPGTLVXTAFGFLRRCVTRGCRVFSDCLYCFARLQAVRDVLHQ
151	IGERCLOSWFAIGIPAARFDYFFDCGNDGRCVCRCEFRCEICRTRLEAD
201	YFPDCGNDGGRCVCRCEFRCEICRTRLEAD
251	RNPYQMIVS MF*

ORF109ng-1 and ORF109-1 show 98.9% identity in 262 aa overlap:

50	orf109ng-1.pep	10	20	30	40	50	60
	orf109ng-1.ng						
55	orf109ng-1.pep	10	20	30	40	50	60
	orf109ng-1.ng						
60	orf109ng-1.pep	70	80	90	100	110	120
	orf109ng-1.ng						

	orf109-1	TVSFARKGLIDWKKGLPIAAASFVGGVAGALSVSLVSKDILLAVVPLLIIFVALVFVFSF	70	80	90	100	110	120
5	orf109ng-1.pep	KLDGSKEGKARMSFFFLGLTVAPLGLGYDGVFGPGVGSFFLIAFIATVLLGCKLLNAMSYTK	130	140	150	160	170	180
	orf109-1	KLDGSKEGKARMSFFFLGLTVAPLGLGYDGVFGPGVGSFFLIAFIATVLLGCKLLNAMSYTK	130	140	150	160	170	180
10	orf109ng-1.pep	LANVACNLGSLSVFLHGSIIFFPIATMAVGAFAVGANLGARFAVRFGSKLIKPLLIIVISI	190	200	210	220	230	240
	orf109-1	LANVACNLGSLSVFLHGSIIFFPIATMAVGAFAVGANLGARFAVRFGSKLIKPLLIIVISI	190	200	210	220	230	240
15	orf109ng-1.pep	SMAVKLLIDERNFLYQMIVSMFX	250	260				
20	orf109-1	SMAVKLLIDERNFLYQMIVSMFX	250	260				

In addition, ORF109ng-1 shows homology to a hypothetical *Pseudomonas* protein:

sp|P29942|YCB9_PSEDE HYPOTHETICAL 27.4 kD PROTEIN IN COBO 3'REGION (ORF9)
>gi|551924|pir|I138164 hypothetical protein 9 - Pseudomonas sp >gi|551924|
25 (M62866) ORF9 [Pseudomonas denitrificans] Length = 261
Score = 175 bits (439), Expect = 3e-43
Identities = 83/214 (38%), Positives = 131/214 (60%), Gaps = 1/214 (0%)
Query: 41 PPVSIAITNKLQXXXXXXXXXXXXXKGLIPKXXXXXXXXXXXXXXXXXXXX 100
PP+ + TNKLG R+G + + K+ L+ D+
Sbjct: 43 PRLCTLTGNKLGLOGFGSGSATLSYARGHVNLEKQLPMAIMSAGAVLGALLATIVPGDV 102
Query: 101 LLAVFVLLIFLVALYFVFSPKLDGSKMNRSMFFGLTIVAPLLGFYDVGFGVGVSGL 160
L A++P LLI +ALYF P + G + +R+ + +F+G L+T+ PL+GFGYDVGFGVG GSF
Sbjct: 103 LKAILFLLIATLALYFGLKNM-GVDVDSRVTFVFTTITLVLPGFYGFGFGTGSF 161
Query: 161 LIAFIVLLGCKLLMAMSYTILANWAGNCGLSIFVFGSLIFIFPVATMVGAVGAVNLGA 220
+ + F + L G +PLA + +TK N N+G+ VFL G++++ + M +G +FA +G+
Sbjct: 162 MLCGFTVLACGVLKATAHTKFLNFGSNMVG+VFLFGGAVMLKVGILLMGLCGQLGAVGS 221
40 Query: 221 RFARVRFSGKILPKLLIVISISMVAVKLLIDERNPL 254
R+A+ G+K+IKPLL++SI++L+L D +PL
Sbjct: 222 RYAKAGAKIIPKLLVIVISIALAIPALADPHT 255

Based on this analysis, including the presence of a putative leader sequence (double-underlined)

45 and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 52

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 441>:

50	1	...	CTCGTAGGGT	ATTGATCTGGC	TATTGGGTGAT	GGCTGTGTGGA	GCAAACATTG
	5		CCGACGACGG	ATATTTTGGT	CAATTTCCGA	TCGTTTGGG	CGAG, ATCTT
	101		TGGTTTCTTG	GGAGTCTATG	AAGTCATATG	TTCGGCATGG	TTTGTCGTTA
	151		TCATGATGTT	TTTGGTGGTT	TCACAGCAT	TGTGGCTGAT	TCGCAATGTG
	201		CCGCGGCTTT	GGCGCGAAT	GAGTCTCTTT	CGGCAAAAGG	TTAAGAAAAA
55	251		ATCTCTGGCG	GACGTGCGCC	ATTCTCTCGT	TGTGGATGTA	AAATTTGCGC
	301		CGAGGATGCG	CAAACTGTT	CTGGAAGATC	AAGGTTTCCA	GGGGAACACC
	351		ATTAACTGGT	AAGACGGGTC	GGTCTGTGAT	CGCGCCAAAA	AAGGCACAAT
	401		GAAACAATGG	GGCTATATCT	TGGCCATCAT	TGCTTTGATT	GTCAITTGATC
	451		GGACGGGGTT	GATAGACAGT	AGGCTCGCT	TGAACATGGT	TATGCTGCC

551 CCGAAAGTAT .TTTGGGTGC gTCCAATCTC TCATTTAGGG GCAACGTCAA
601 TATTTCGG.A GGGGCAGgT GCGGATGTGG TTTTCCTGA

This corresponds to the amino acid sequence <SEQ ID 442; ORF110>:

5 1 ..LLGIASVIGT LLQONQPQTD YLVKFGSFWA XIFGFLGLYD VYASAWFVVI
51 MMFLVVSSTL CLIRNVPPFW REMKSFREKV KEKSLAAMRH SSLLDVKIAP
101 EVAKRYLEVQ GFQGKTINRE DGSVLIAAKK GTMNKNGYIF AHVALIVICL
151 GGLIDSNLLL KLGMLTGRIF RTIRREMPRI XKPESXFGCV QSLI*GQRQY
201 FXRGRVMMWF S*

Computer analysis of this amino acid sequence gave the following results:

10 Homology with ORF88a from *N.meningitidis* (strain A)

ORF110 shows 91.5% identity over a 188aa overlap with ORF88a from strain A of *N. meningitidis*:

		10	20	30	40	50	60
orf88a.pep	MSKSRSPPLSRPFAFFSSMRFAVALLSLG	IASVIGT	VLQONQPQTD	YLVKFGSFWA			
15 orf110				LLG	IASVIGT	VLQONQPQTD	YLVKFGSFWA
					10	20	30
		70	80	90	100	110	120
20 orf88a.pep	QIFGFLGLYDVYASAWFVIMFLVVSSTLCLIRNVPPFWREMKSFREKVKEKSLAAMRH						
orf110	XIFGFLGLYDVYASAWFVIMFLVVSSTLCLIRNVPPFWREMKSFREKVKEKSLAAMRH						
		40	50	60	70	80	90
		130	140	150	160	170	180
25 orf88a.pep	SSLLDVKIAPEVAKRYLEVQGFQGKTINREDGSLVIAAKGTMNKNGYIFAHVALIVICL						
orf110	SSLLDVKIAPEVAKRYLEVQGFQGKTINREDGSLVIAAKGTMNKNGYIFAHVALIVICL						
		100	110	120	130	140	150
		190	200	210	220	230	240
30 orf88a.pep	GGLIDSNLLLKLGMLTGRIVPDNQAVYAKDFKPESILGASNLSPFRGNVISEGQSAUVVF						
orf110	GGLIDSNLLLKLGMLTGRIFRTIRREMPRIKXKPFSGVQSLIXGQRQYFXRGRVMMWF						
35 orf88a.pep	LNAENGILVQDLPEFVKLKKFHDIFYNTGMPERDFASDIEVTDKATGEKLEKRTIRVNHPLT						
orf110	SK						

40 However, ORF88 and ORF110 do not align, because they represent two different fragments of the same protein.

Homology with a predicted ORF from *N.gonorrhoeae*

ORF110 shows 88.6% identity over a 211aa overlap with a predicted ORF (ORF110.ng) from *N. gonorrhoeae*:

45 orf110.pep		LLGIASVIGTLLQONQPQTDYLVKFGSFWA	30
orf110ng	MSKSRISPTLLSRPFAFFSSMRFAVALLSLG	IASVIGTLLQONQPQTDYLVKFGFPWT	60
orf110.pep	XIFGFLGLYDVYASAWFVIMFLVVSSTLCLIRNVPPFWREMKSFREKVKEKSLAAMRH		90
orf110ng	RTDFGLGLYDVYASAWFVIMFLVVSSTLCLIRNVPPFWREMKSFREKVKEKSLAAMRH		120
orf110.pep	SSLLDVKIAPEVAKRYLEVQGFQGKTINREDGSLVIAAKGTMNKNGYIFAHVALIVICL		150
55 orf110ng	SSLLDVKIAPEVAKRYLEVQGFQGKTINREDGSLVIAAKGTMNKNGYIXAHVALIVICL		180


```

10      1  MSKSRISPTL  LSRPWFAPFS  SMRFAVALLS  LLGIASVIGT  VLQQNQPTD
51     51  YLVKFGPFWT  RIDFDGLGLY  VYASAVDVAT  MMFLVVSSTL  CLIRNVPSPV
101    101  REMKSFREKV  KFKSLAAMRH  SLLLDVFKPI  EVAKRYLEVR  GFQKGTVSRV
151    151  DGSVLIAAKK  GKMSKWGXYI  AHVALITVCL  GRLINXNLLL  KGLMGLSGIF
201    201  RNRBRVMPRT  SPKNSTGWGV  OST.IKGOROV  FORGKRWMMF  S*

```

Example 53

20	1	ATCGCGGCTG	AAACGAGCCT	GGCGAATCTG	ATCCGGCTGT	TGATATTTCG
	5	CTCGGGTCTT	ATCTCTCTGA	CGGCTGTCTC	GGACAACCC	GGCAACACCG
	11	TACCCCTGTC	GGGCGAAACG	GGGCGACGGA	CGATACCGT	CAAAATACCTT
	15	TCAATATATC	GGGCAAACT	CGCCTACCT	CCGCAAAATC	AAAAACGATC
	20	CGATGACGGC	TTCTAAGAAG	CAACACGGCA	GATGTCCAC	TATGACGCCG
25	25	ATCTCGAAT	CAGCGGTT	TACCAACGCA	CAGCGCCGAC	GGCCCTCCGC
	30	ATTCTCAGCG	ATCTGCGACA	CGTTATCGCG	GAAAGCGCTC	GGCTGAACCG
	35	CTGTACACAC	GGGCGGGTGG	AGGTAACGCT	GTGACCGTCT	GTGACACTTT
	40	GGGGAATCGG	CGGCGAAGAA	CGGTTACCC	CGGCAAAACG	GGCGATCAAT
	45	ATCAAAATAT	CGGCTCTT	TAGGACGCA	CGGCAAAACG	TTTGGAATAT
	50	AGGCAAAAT	TACGCTCTCT	TGGCAAGAAC	CGACCCCAAG	GGCTATTTCG
	55	ATTCTATCT	GATTGCCAAA	GGTCTCGCG	TGTGATAAT	TGCGGGCGAA
	60	CTGGAAATAT	AGCGCATTTCA	AAATTTATCTG	TCGGCGAGTT	GGCGCGAGTT
	65	GACGCGGCAAA	GGCAAAAACG	CGCGCGCGGA	ACCGTGGCG	ATCGGTATCTG
	70	AGGAGCGCCAA	TATCGTCCGCA	GGCGGCAAT	CGGAGATAT	GTCTCCGATCG
35	75	AAACACCGTT	CGCTGTGCAC	TCTCGGCGAT	TACCGTTATT	TCCAGCTCGA
	80	TAAAACCGG	AAAGCGCTCT	CCGATATCAT	TACCGGATAC	AACAACGACG
	85	CGATGACGCA	CAACCTGGCC	TCACTCAGCG	TGGTGGCGCA	CAGTGGCGAG
	90	ACGGGCGGCG	CTGTTCGAC	AGGATATCT	GTAATGGGCG	AAACCGAAGT
	95	CTTAAAGCG	CTAGGCGCG	AAAGACGCG	ATTCGCGGCG	ATTCGCGGCG
40	100	ATGAAGCGG	CTACGCGCAC	GGCAATCTT	CGGAAATGGA	AAAACCTGTC
	105	CTGATA				

	1	MPSETRLPNF	IRVLILFALGF	IFLNVASQET	AGQTVLQGET	MGTTYTVKYL
45	51	SNRKEKLPSP	AEIQKRIIDA	LKEVNRQMSI	APQSTLSRF	NQHTGKPLR
	101	TSDFAHVY	EAVILNRILTH	DGLLVGTGSL	AVYGLSSG	SVFTGPKVGR
	151	IQKASQYTYI	DKILLKQGLD	YASLSTHPK	YVLEKSLD	SVFTGPKVGR
	201	LEKYGIQNYI	VIIGELHGLH	GRNANGPWR	IGKQIPNVQ	GGNTQIVLPL
	251	NRRSLATSGD	YRIELHDKNG	KRILSHINPL	NRKRISHMLA	SISVADSAM
50	301	TADGLSTGLF	VLGETEALKL	AEREKLAFLV	IVRDKGGYRT	AMSSEPEKLL
	351	R*				

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N. meningitidis* (strain A)

ORF111 shows 96.9% identity over a 351aa overlap with an ORF (ORF111a) from strain A of *N. meningitidis*:

		10	20	30	40	50	60
5	orf111a.pep	MPSETRLPNFI	IRTLIFALSF	IFLNACSEQT	AQTVTLQGET	MGTTYTVKYL	SNNRDXLPSP
	orf111	MPSETRLPNFI	IRVLIFALGFI	IFLNACSEQT	AQTVTLQGET	MGTTYTVKYL	SNNRDXLPSP
		10	20	30	40	50	60
10	orf111a.pep	AEIQXRIDDAL	KEVNRQMS	TYQPDSEISR	FNQHTAGK	PLRISSDFA	HTAEAVHLNRLTH
	orf111	AEIQXRIDDAL	KEVNRQMS	TYQPDSEISR	FNQHTAGK	PLRISSDFA	HTAEAVHLNRLTH
		70	80	90	100	110	120
15	orf111a.pep	GALDVTVGFL	VNLWGF	PGDKSV	TRPSP	EQIKQAA	SYTGIDKII
	orf111	GALDVTVGFL	VNLWGF	PGDKSV	TRPSP	EQIKQAA	SYTGIDKII
		130	140	150	160	170	180
20	orf111a.pep	AYLDLS	SSIAGK	FGVDK	VAGLEK	YGIQNYL	VEIGGELH
	orf111	AYLDLS	SSIAGK	FGVDK	VAGLEK	YGIQNYL	VEIGGELH
		190	200	210	220	230	240
25	orf111a.pep	GGNTQI	IVPLN	RSXAT	SGDYR	IFHVD	SGKRLSHI
	orf111	GGNTQI	IVPLN	RSXAT	SGDYR	IFHVD	SGKRLSHI
		250	260	270	280	290	300
30	orf111a.pep	TADGXSTGL	FVLG	SETALK	LAEREK	IAVFL	IVRD
	orf111	TADGXSTGL	FVLG	SETALK	LAEREK	IAVFL	IVRD
		310	320	330	340	350	
35	orf111a.pep	CGGCGG	CGA	CAACTTC	ATCGCG	CACTT	TGATATT
	orf111	CGGCGG	CGA	CAACTTC	ATCGCG	CACTT	TGATATT

The complete length ORF111a nucleotide sequence <SEQ ID 447> is:

40	1	ATGCCGCTCTG	AAACACGCGCT	GCGGAACCTT	ATCGCGCACTT	TGATATTGTC
	51	CCTGAGTTT	ATCTCTCTGA	ACGCTCTGTC	GGAACAAACC	GCGCAAAACC
	101	TTACCCCTGCA	AGGTGAAACG	ATGCGCACGA	CCTATACCGT	CAAAATACCTT
	151	TCGAATAATC	GGGACNAACT	CCGNTACCTT	GCCGAATAAT	AAAACGCAAT
45	201	CGATGACGG	CTTAAAGAC	TACACCGCA	GATCTCCACC	TATCAGCCG
	251	ACTCCGAAT	CAGCGGTTT	AACCAACACA	CAGCGGCAA	GCCTCTCCGC
	301	ATTTCACGC	ACTTCGCACA	CGTTACTGCC	GAGCGCTCC	ACCTGAACGC
	351	CCTGACACAC	GCGCGCGTGG	ACGTAACCGT	CGGCCCCCTG	GTCACCTTT
	401	GGGGATTGCG	CCCCGACAAA	TCCGTTACCC	GTGAACCGTC	GCCGGAACAA
	451	ATCAAAACAG	CAGCATCTTA	TACGGGCATA	GACAAAATCA	TTTTGAACAA
50	501	AGGCAAGAT	TACGCTTCCT	TAGCAAAAAC	CCACCCCAAG	GCCTATTGTC
	551	ATTATCTTC	GATTGCCAAA	GGCTTCGGCG	TTGATNAGT	TGCGGCGGAA
	601	CTGGAAAAAT	ACGGCATCTA	AAATTTATCT	GTGCAATCG	GCGGAGAGTT
	651	GCAAGCGCAA	GNCAAAAACG	CGCGCGGCGA	ACCTTGGCGC	ATCGGCATCG
	701	AACAGCCCAA	CATCGTCCAA	GCGCGCAATA	CGCAGATTAT	CGTCCCGCTG
55	751	AACAAACGTT	CNNTTGCAC	TTCCGCGCAT	TACCGTATT	TCCAGCTCGA
	801	TAAAGCGCG	AAACGCTCT	CCCATATCAT	TAATCCGAC	ACCAACGCAC
	851	CCATGACCA	CACTCCCG	TCATCAGCG	TGNTCCGAGA	CAGTGCGAT
	901	ACGCGGACG	CGTNTCCAC	AGGATATTC	GTATTGGCG	AAACCGAAGC
	951	CTTAAAGCTG	GCAGAGCCCG	AAAACTCGC	TGTTTTCCTG	ATTGTCAGGG
60	1001	ATAAAGCGG	CTACCGCAC	GCCATTCCTT	CCGAATTGTA	AAAACCTGCT
	1051	CGCTAA				

This encodes a protein having amino acid sequence <SEQ ID 448>:

```

1  MPSETRLPNF IRTLIFALSF IFLNACSEQT AQTVTLQGET MGTTYTVKYL
51  SNNRDXLPSP AEIQXRIDDA LKEVNRQMS  TYQPDSEISR FNQHTAGKPLR

```

101 ISSDFAHVTA EAVHINRLTH GALDVTVGPL VNLWGFDPK SVTREFSPQ
 151 IKQAASTYGI DKIIILKQGD YASLSKTHPK AYLDLSSIAK GFGVDXVAGE
 201 LEKYGIQNYL VEIGGELHGK XKNARGEPR IGIEQPNIVQ GNTQIIVPL
 251 NNRSKATSGD YRIFHVDPKSG KRLSHINPN NKRPISHNLA SISVXADSAM
 301 TADGXSTGLF VLGETEALKL AEREKLAVFL IVRDKGGYRT AMSSEFEKLL
 351 R*

Homology with a predicted ORF from *N.gonorrhoeae*

ORF111 shows 96.6% identity over a 351aa overlap with a predicted ORF (ORF111.ng) from *N.*

10 *gonorrhoeae*:

		10	20	30	40	50	60
orf111ng		MPSETRLPLNLRALIFALGFIFLNACSEQAQTVTLQGETMGTTYTVKYLNNRDKLPSP					
orf111		MPSETRLPLNLRALIFALGFIFLNACSEQAQTVTLQGETMGTTYTVKYLNNRDKLPSP					
		10	20	30	40	50	60
		70	80	90	100	110	120
orf111		AKIQKRIDDALKEVNRQSTYTQDSEISRFNQHTAGKPLRISDDFAHVTAEAVRINRLTH					
orf111		AEIQKRIDDALKEVNRQSTYTQDSEISRFNQHTAGKPLRISDDFAHVTAEAVRINRLTH					
		70	80	90	100	110	120
		130	140	150	160	170	180
orf111ng		GALDVTVGPLVNLWGFDPKSVTREFSPQIKQAASTYTGDKIILQQKGDYASLSKTHPK					
orf111		GALDVTVGPLVNLWGFDPKSVTREFSPQIKQAASTYTGDKIILQQKGDYASLSKTHPK					
		130	140	150	160	170	180
		190	200	210	220	230	240
orf111ng		AYLDLSSIAKGFVDPKVALEKYGIQNYLVEIGSELHGKGNHAGEFWRIGIEQPNIIQ					
orf111		AYLDLSSIAKGFVDPKVALEKYGIQNYLVEIGSELHGKGNHAGEFWRIGIEQPNIIQ					
		190	200	210	220	230	240
		250	260	270	280	290	300
orf111ng		GGNTQIIVPLNNRSLATSGDYRIHVDPKGNKRLSHIINPNKRPISHNLASISVVDSDAM					
orf111		GGNTQIIVPLNNRSLATSGDYRIHVDPKGNKRLSHIINPNKRPISHNLASISVVDSDAM					
		250	260	270	280	290	300
		310	320	330	340	350	
orf111ng		TADGLSTGLFVLGETEALKLAREKLAVFLIVRDKGGYRTAMSSEFAKLLRX					
orf111		TADGLSTGLFVLGETEALKLAREKLAVFLIVRDKGGYRTAMSSEFAKLLRX					
		310	320	330	340	350	

The complete length ORF111ng nucleotide sequence <SEQ ID 449> is:

1 ATGCCGCTCTG AAGACAGCCT GCGGAACCTT ATCCGGCCTT TGATATTTGC
 51 CCTGGGTTTC ATCTCTCTGA ACSCCTGTTC GGAACAACCC GCGCAACCCG
 101 TTACCCCTGCA AGCGCAAAAG aTGGGTACGA CCTATACCGT CAATACTCTT
 151 TCATAATATC GGGCAAAACT CCCCCTCCCT GCCAAAATAC AAAAGCGCAT
 201 TGATGATGCG CTTAAAGAAAG TCAACCGSCA GATGTCCAC TACCAGACCG
 251 ATTCGGAAT CAGCGGTTC AACCAACACA CAGCCGGCAA GCCCTCCCGC
 301 ATTTCAAGCG ATTTGCGACA CSTTACCGCC GAAGCGGTCC GCTGAACCG
 351 CCTGACTCAC GGCSCACTG ACSTAACCGT CGGCCCTTTC GTCAACCTTT
 401 GGGGTTTCGG CCCCAGACAA TCCGTTACCC GTGACCGCTC GCGGAACAA
 451 ATCAACAGG CGGCATCTTA TACGGGATA GACAAATCA TTTTGACACA
 501 AGCAGAGAT TACGCTTCTT TGACAAACG CACCCCAACA GCCTATTGCG
 551 ATTATCTTTC GATTGCAAAA GGCCTCGGGC TTGATAAAGT TGGGCGCAA
 601 CTGGAATAAT ACSCGATTC AATATTCTG GTCGAAATCG cgggcGAGTT
 651 GCACGCGAAA GGCAGAAATG CGCACGCGCA ACCGTGGCGC ATCGGTATAG
 701 AGCAACCCAA TATCATCCAA GgcgcGCaata CGCAGATTAT cgtcccgctg
 751 aaCaacogtt cgtTGCCAC TTCCGGCGAT TaccgtaTTT tccacgtcgA
 801 TAAAAcggc aaacgccttt cccacaTCAT CAATCCCAcA aacAAcGcag
 851 ccATCAGcca caacctcgcc tccatcagcg tggctcGAG CAGTGCATG
 901 ACGGCGGACG GTTtatCCAC AGGATTATT GTTTTACGCG AAACCGAAGC
 951 CTTAAGGCTG GCAGAACCAAG AAAAATTCGC TGTTTTCCTA ATTGTCGCGG

1001 ATAAGGACGG CTACCGCACC GCCATGTCTT CGAATTTCG CAAGCTGCTC
1051 CGCTAA

This encodes a protein having amino acid sequence <SEQ ID 450>:

5 1 MPSETRLPNL IRLALIFALGF IFLNACSEQT AQTVTILOGET MGTTYTVKYL
51 SNNRDKLPSP AKIQKRIDDA LKEVNRQMS YQTDSEISRF NQHTAGKPLR
101 ISSDFAHVTA EAVRLNRLTH GALDVTVGPL VNLWGFSGPK SVTREPSPQ
151 IKQAASTYGI DKIIILQQQKD YASLSKTHFK AYLDLSSIAK GFGVDKVAE
201 LEKYGIQNYL VEIGGELHKG KGNARSGPWR IGIEQPNIIQ GGNTQIIVPL
251 NNRSLATSGD YRIFHVDKNG KRLSHIINFN NKRPISHLNA SISVSDSAM
10 TADGLSTGLF VLGETEALRL AEQEKLAFLV IVRKDKGYRT AMSSEFAKLL
351 R*

This protein shows homology with a hypothetical lipoprotein precursor from *H. influenzae*:

sp|P44550|YOJL_HAFLIN HYPOTHETICAL LIPOPROTEIN HI0172 PRECURSOR >gi|1074292|pir|4
15 hypothetical protein HI0172 - Haemophilus influenzae (strain RD KW20)
>gi|1573128 (U32702) hypothetical [Haemophilus influenzae] Length = 346
Score = 353 bits (896), Expect = 9e-97
Identities = 181/344 (52%), Positives = 247/344 (71%), Gaps = 4/344 (1%)
Query: 7 LPNLIRALIFALGFIFLNACSEQTAVTILQGETMGTTYTVKYLNNRDKLPSPAKIQKR 66
+ LI +I + L AC +T + ++L G+TMGTTY VKYL + S K +
Sbjct: 1 MKKLISGIIAIVAMALSLAACQKET-KVISLSGKTMGTTYHVKYLDGGSITATSE-KTHEE 58
Query: 67 IDDLKEVNRQMSYQTDSEISRFNQHT-AGKPLRISDDFAHVTA EAVRLNRLTHGALDV 125
I+ LK+VN +MSTY+ DSE+SRFQ+T P+ IS+DFA V AEA+RLN+T+T GALDV
25 Sbjct: 59 IEAILKOVNMAMSTYKDKSELSRFNQHTQVNTPIEISADPAKVLAEAIRLNKVTGALDV 118
Query: 126 TVGFLVNLWGFSGPKSVTREPSPQIKQAASTYGDIDKILQQGKDYASLSKTHPKAYLDL 185
TVGP+VNLWGFSGP+K ++P+PEQ+ + ++ GIDKI L K+ A+LSK P+ Y+DL
Sbjct: 119 TVGFPVNLWGFSGPEKRPKQPTPEQLAERQAWVGIDKITLDNKEKATLSKALPQVYVDL 178
30 Query: 186 SSIAGKFGVDKVAELEYGIQNYLVEIGGELHKGKGNARSGPWRIGIEQPNIIQGGNTQ 245
SSIAGKFGVD+VA +LE+ QNY+VEIGE+ KGNK G+PW+I IE+P +
Sbjct: 179 SSIAGKFGVDQVAEKLQLNAQNYMVEIGGEIRAKGNIEGKFWQIAIEKPTTTGRAVE 238
35 Query: 246 IIVPLNNRSLATSGDYRIFHVDKNGKRLSHIINFNNKRPISHNLASISVSDSAMTADGL 305
++ LNN +A+SGDYRI+ ++NGKR +H I+P PI H+LASIAV++ +NTADGL
Sbjct: 239 AVIGLNNMGMASSGDIYRIY-FEENGKRPFAHEIDPKTGYPQIHHLASITVLAPTSMTADGL 297
40 Query: 306 STGLFVLGETEALRLAEQEKLAFLVIRVDKGYRTAMSSSEFAK 349
STGLFVLGE +AL +AE+ LAV+LI+R +G+ T S S F KL
Sbjct: 298 STGLFVLGEDKALEVAEKNNLAVYLIIRTDNGFVTKSSSAFKKL 341

Based on this analysis, it is predicted that the proteins from *N. meningitidis* and *N. gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 54

45 The following partial DNA sequence was identified in *N. meningitidis* <SEQ ID 451>:

1 ..CGGTGCCGCC GACAGGCGCA CGAGCTGTAT GCGGCGCAGC CGTCCCGTCA
51 AAAATTTGTTG CTGCGCTTCA TCGGCGCGCG GTGCGCATCA AATATACGGC
101 GCGGCGCGCG TCGCGACGCG TCGCGCAAAAG GCGTGCMAAT CGGCGCGCAG
50 151 GTGTTTGTAT GGCACAAATGA AGGCAGCCCA yTGCCAAATCG CGGTGATGGG
201 CGGCGAGGCGC GGCCAGCACG CWTACGTCAA CGGCCAAGGCG GGTGCGGCAG
251 gCAGTGATT TTATGCTTAT GgCGGGGgTg TTTATGCTgC GTGGCATCCG
301 TTGCGCGATA AACAAACGGG TgCGTATTG GACGCGTGTG TGCAATACCA
351 ACGTTTCAAA CACCGCATCA ATGATGAAAA CGGTGGGGA CgCTACAAAA
401 CCAAGGTTG GACGCGCTTC GTGCAAGGCG GCTACAAAGC GCTTGTGGCG
55 451 GAAGGCATTG TCGGAAAGG CAATAATGTG CGGTTTTRAC TACAACCGCA
501 GgCGCAGTTT ACCTACTTGG CGATAACGG CGGCTTTRAC GACAGCGAGG
551 GGAGCGCGGT CGGACTGCTC GGCAGCGGTC AGTGGCAAG CCGCGCGGCG
601 ATTCGGGCAA AAGCCGTTT TCGTTTGGCT AACGCTGCT ATCTTCAGCC
651 TTTTCCGCT TTTAATGTT TGCACAGTC AATATCTTC GGCGTGA
60 TGGACGGCGA AAAACAGAGC CTGCGAGGCA GGACGCGACT CGAAGGCGCG

-271-

751 TTCGGTATTG AAGCCGGTTG GAAAGGCCAT ATGTCOCCA..

This corresponds to the amino acid sequence <SEQ ID 452; ORF35>:

```

1  ..PCRRQGDVVY AAHASRQKLW LRFIGGRSHQ NIRGGAAADG WRKGQVIGGE
51 VVFRQNEGSX LAIGVMGGRA QHASVNGKG GAAGSDLYGY GGGVYAAWHQ
101 LRDKQTGAYL DGWLQYQRFK HRINDENRAE RYTKGWTAS VEGGYNALVA
151 EGIVGKGNV RYFLQPOAQF TYLGVNGGFT DSEGTAVGLL GSGQWQSRAG
201 TRAKTRFALR NGVNLQPFPA FNVLHRSKSF GVMDEGKQT LAGRTALEGR
251 FG:EAGWKGH MSA..

```

Computer analysis of this amino acid sequence gave the following results:

10 Homology with putative secreted VirG-homologue of *N. meningitidis* (accession number A32247)

ORF and virg-h protein show 51% aa identity in 261aa overlap:

```

Orf35 5 QGDDVYAAHASRQKLWLRFIGGRSHQNIRGGAA-ADGWRKGVQIGGEVFRQNEGSXLA 63
+ D++ R+ LMLR I G S+Q ++G A +G+RKGQV+GGEVF QNE + L+I
virg-h 396 KNSDI FDRTLPRKGLWLRVIDGHSNQWVGKTA PVGSEYKGVGLGGEVFTWQNE SNLSI 455

Orf35 64 GVMGGRAGQHASVNGKG--GAAGSDLYGYGGGVYAAWHQLRDKQTGAYLDGWLQYQRFKH 121
G+MGG+A Q ++ + ++ G-G GVYA WHQL+DKQTGAY D W+QYQRF+H
virg-h 456 GLMGGQAQRSTFHNPDNDLTGNVKGFGAGVYATWHQLDKQTGAYADSWMQYQRFH 515

Orf35 122 RINDENRAERYKTKGWTASVEGGYNALVAEIVGKGNVFRYLPQQAQFTYLGVNGGFTD 181
RIN E+ ER+ +KG TAS+E GYNAL-AE KGN++R YLPQQAQ TYLGVNG F+D
virg-h 516 RINTEDGTERFTSKGITASIEAGYNALAEHFTKGNLSRVLYLPQQAQFTYLGVNGGFTD 575

Orf35 182 SEGTAVGLLGSQWQSRAGIRAKTRFALRNGVNLQPFPAFNVLHRSKSPGVMDEGKQTL 241
SE V LLGS Q Q+R G++AK +F+L + ++PFPA N L+ +K FGVMDEG++ +
virg-h 576 SENAHVNLGSRQLQTRVGVQAQAQFSLYKNATIEPFPAFVNALYHNKFPFGVMDEGERRVI 635

Orf35 242 AGRTALEGRPGIEAGWKGHMS 262
+TAE + G+ K H++
virg-h 636 NNKTAIESQLGVAVKIKSHLT 656

```

Homology with a predicted ORF from *N. meningitidis* (strain A)ORF35 shows 96.9% identity over a 259aa overlap with an ORF (ORF35a) from strain A of *N.**meningitidis*:

```

35 orf35.pep 10 20 30
PCRRQGDVVYAAHASRQKLWLRFIGGRSHQNTIRG
orf35a 310 320 330 340 350 360
QRLAIPEAEAVLYAQQAYAANTLFGI RAADRGDDVYAADPSRQKLWLRFIGGRSHQNTIRG

40 orf35.pep 40 50 60 70 80 90
GAAADGWRKGVQIGGEVFRQNEGSXLAIGVMGGRAGQHASVNGKGGAAGSDLYGYGGGV
orf35a 370 380 390 400 410 420
GAAADGRRKGVQIGGEVFRQNEGSRLAIGVMGGRAGQHASVNGKGGAAGSYLHGYGGGV

45 orf35.pep 100 110 120 130 140 150
YAAWHQLRDKQTGAYLDGWLQYQRFKHRINDENRAERYKTKGWTASVEGGYNALVAEGIV
orf35a 430 440 450 460 470 480
YAAWHQLRDKQTGAYLDGWLQYQRFKHRINDENRAERYKTKGWTASVEGGYNALVAEGVV

50 orf35.pep 160 170 180 190 200 210
GKGNVNRVLYLPQQAQFTYLGVNGGFTDSEGTAVGLLGSQWQSRAGIRAKTRFALRNGVN
orf35a 490 500 510 520 530 540
GKGNVNRVLYLPQQAQFTYLGVNGGFTDSEGTAVGLLGSQWQSRAGIRAKTRFALRNGVN

55 orf35.pep 220 230 240 250 260
LQPFPAFNVLHRSKSPGVMDEGKQTLAGRTALEGRPGIEAGWKGHMSA
orf35a 550 560 570 580 590
LQPFPAFNVLHRSKSPGVMDEGKQTLAGRTALEGRPGIEAGWKGHMSA

```

-272-

orf35a LQPFPAFNVILHRSKSFVEMDGEKOTLAGRTALEGRFIEAGWKGHMSARIGYKRTDGD
550 560 570 580 590 600

orf35a KEAALSLKWLFX
610 620

The complete length ORF35a nucleotide sequence <SEQ ID 453> is:

1 ATGTTCCAGAG CTCAGCTTGG TTCAAATACT CGTTCTACCA AAATCGGCGA
51 CGATGCCGAT TTTTCAATTT CAGACAAGCC GAAACCGGCG ACTTCCCAAT
101 ATTTTTCAG CGGTAAAAAC GATCAAAAT CATCOGAAT TGGGTATGAC
151 GAAATCAATA TCCAGGGTAA AACTACAAT AGCGGCATAC TCGCGTGTGA
201 TAATATGCC GTTGTTAAGA AATATATTAC AGATACTTAC GGGGATAATT
251 TAAAGGATGC GTTAAAGAG CAATTACAGG ATTTATACAA AACAGAGCCC
301 GAAGCTTGGG AAGAAAAATA AAAACGGACT GAGGAGGCGT ATATAGAACCA
351 GCTTGGACCA AAATTTAGTA TACTCAACA GAAAACCCCG GATTTAATTA
401 ATAAATGSGT AGAAGATTCC GTACTCACTC CTCATAGTAA TACATCACAG
451 ACTAGTCTCA ACAACATCTT CAATAAAAAA TTACACGTCA AAATCGAAAA
501 CAAATCCAC GTCCCGGAG AGGTTGTGGA ACTGACCAAG ATGACGCTGA
551 AAGATTCCCT TTGGGAACCG CGCGCCATT CGCATACCA TATGCTGGAA
601 ACTTCCGATA ATGCCCGCAT CGCGCTGAAC ACGAAGATGT AAAAATGACG
651 CGTCCATAAA GCGTATCAGG GCGGTGCGGA TTTCCTGTTC GGCTACAGAG
701 TGCGGGAGTC GGACAAACCC GCGCTGACCT TTGAAGAAAA AGTCAGCGGA
751 CAATCCGCGG TGGTTTTGGA ACGCCGCGCG GAAATCTGA AAACGCTCGA
801 CGGCGCAAAA CTGATTTCGCG CCGAAAAGCG AGACTCTAAT TCGTTTGGCT
851 TAAACAAA GTTACGCGAG GACATGTAGC AATTATTGCT CAACGATCG
901 GAAGCGAGAT TTTGCTTGGG CGTGACGCGT TTGGCTATCC CCGAGCGCGA
951 AGCGGTTTTA TATGCCAAC AGGCTATGCT GGCAATACT TTGTTGCGGG
1001 TCGCTGCCGC CGACAGGGGG GACGACGTGT ATGCGCGCGA TCCGTCGCT
1051 CAAAATTTGT GSGTCGCTT CATCGCGCGG CGGTCGCATC AAAATATACG
1101 GGGCGCGCGC GCTGCGGAGC GGGCGCGCAA AGGCGTGCAA ATCGCGCGCG
1151 AGGTGTTTGT ACGGCAAAAT GAAGGACGCG GGCTGCGAAT CGGCGTATG
1201 GGGCGCAGGG CTGCGCAGCA CGCATCAGTC AACGGCAAG GCGGTGCGCG
1251 AGGCAGTTAT TTGCTAGTGT ATGGCGGGGG TGTATTGCT GCGTGGCATC
1301 AGTTGCGCGA TAAACAAAC GGTGCGTATT TGGACGCTGT GTTGCAATAC
1351 CAACGTTTTA AACACCGCAT CAATGATGAA AACGTCGCGG AACGCTACAA
1401 AACCAAAGGT TGGACGGCTT CTGTCGAAGG CGGCTACAC CGGTTGTGG
1451 CGAAGGCGCT GTTCCGAAA GGCATTAAT GCGCATATG TCGGTTTTA
1501 CAGGCGCAGT TCGCTACTT GCGCTAAAC GCGGCTTTA CCGACAGCA
1551 GGGGACGCGG CTGCGACTGC TCGGACGCGG TCAATGGCAA AGCGCGCGCG
1601 GCATTCGGGG AAAAACCCGT TTGCTTTTGC GTAACGCTGT CAATCTTCAG
1651 CCTTTTGGCG CTTTAAATGT TTTGCACAGG TCAAAATCTT TCGGCGTGGA
1701 AATGGACGCG GAAAAACAGA CGCTGGCAGG CAGGACGCGC CTCGAAGGGC
1751 GGTTCGGCAT TGAAGCGGTT TGGAAAGGCC ATATGTCGCG ACGCATCGGA
1801 TACGGCAAAA GGACGCGACG CGACAAAGAA GCCGCAATGT CGCTCAAAATG
1851 GCTGTTTTGA

45 This encodes a protein having amino acid sequence <SEQ ID 454>:

1 MFAQLGSNT RSTKIGDDAD FFSKDKPKG TSHYFSSGKT DONSSEYGYD
51 EINIQGRNRY SEYLAVDNMP VVKYITDTY GDNLDKAVKQ QLDLYKTRP
101 EAWENKRRK EAYVLEQLG KFSILKQRP DLINKLVEDS VITPHSNTSQ
151 TSLNNIPNKK LHVKIKNSH VAGVLELTK MTLKSLWEP RHSHDHLME
201 TSDNARIRLN TKDKLTVHK AYGGADFLF GYDVRSDGKT ALTPEEKVYG
251 QSGVILERRP ENLKITLGRK LIAAEKADSN SFAFKQNYRQ GLVELLIKQC
301 EGGFCLGVOR LAIPEAEAVL YAOQYAANT LFLRAADRQ DDVYAADPSR
351 QKWLRLFIGG RSHQNIIRGA ADGRRKGVQ IGGEVFRQN EGSRLAIGVM
401 GGRAGOHASV NGKGGAAGSY LHGYGGVGYA AWHQLRDKOT GAYLDGWLQY
451 QRFKRIINDE NRAERYKTKG WTASVEGGYN ALVAEGVVGK GNNVRFYLPQ
501 QAQFTYLVGN GGGTDSGTA VGLLGSQGWQ SRAGIRAKTR FALRNGVNLQ
551 PFPAFNVILHR SKSFVEMDGE EKOTLAGRTA LEGRFIEAG WKGHMSARIG
601 YGKRTDGDKE AALSLKWLFX*

Homology with a predicted ORF from *N.gonorrhoeae*

60 ORF35 shows 51.7% identity over a 261aa overlap with a predicted ORF (ORF35ngh) from *N.*

gonorrhoeae:

orf35.pep PCRRGGDDVYAAHASRQKWLRLFFIGGRSHQNIIRG 34
:::|:: |::|::| |::|::|::|
orf35ngh FTKQERDIDATYAAQAQANTLFALRLNDKNSDIFDRTLPRKGLWRLVIGHNSQWVQG 370

-274-

451 AAAGCGCTTG CCCAAATAT CCGCCTCAAC CTGACCGACA ACCGCAGCAC
501 CGGACACGG CTTCGCCACC GTTTCACAAA TGCCGGTAGT ATGCTGACGC
551 AAGGAGTAGG CGACGGATT CAAACGGCCA CCGGATACAG CCCCAGCTG
601 GCAGATCGG GCAATGCCGC CGAAGCCTTC AACGGACACT GAGATATCGT
651 TAAAAACATC ATCGGCGCTG CAGGAGAAT TGT

This corresponds to the amino acid sequence <SEQ ID 460; ORF46-1>:

1 ..AVCLEPMHAHA SXLANDSFIR QVLDRQHFEF DGKYLHFGSR GELAEQRSHI
51 GLGRTIQSHQL GNLMIQQAAT KGNIGYIVRF SDBGHEVHSP FUNHASSHSD
101 DEAGSFVDFG SLRYIHWDGY EHHPADGYDG PQGGGYPAFK GARDIYSYDI
151 KGVAQNRLIN LTUNRTGQR LAURFHNAGS MLTGQGVGDGF KRATRYSEFL
201 DRSGNAAEAF NQTADIVKNI IGAAAGEI

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N. gonorrhoeae*

ORF46 shows 98.2% identity over a 111aa overlap with a predicted ORF (ORF46ng) from *N.*

15 *gonorrhoeae*:

orf46.pep AEYVQFSIDLFSVSGSGGGIPKAKPVFDAKPRWEVDRLKLNLTTR 45
orf46ng PKTGVPFDGKGFPNFKEHKVYDTKLDIQELSGGGIPKAKPVFDAKPRWEVDRLKLNLTTR 217
20 orf46.pep EQVEKNVQETRRRSQSSQFKAHAQREWENKTGLDFNHFIGGDINKKGTVTGGHSLTRGDV 105
orf46ng EQVEKNVQETRRRSQSSQFKAHAQREWENKTGLDFNHFIGGDINKKGAVTGGHSLTRGDV 277
25 orf46.pep RVIQQTAPDKHGKXLSDDSGN 126
orf46ng RVIQQTAPDKHGVLSDDSGN 298

A partial ORF46ng nucleotide sequence <SEQ ID 461> is predicted to encode a protein having partial amino acid sequence <SEQ ID 462>:

30 1 ..RLKHCCHCAR LGSAFHRQD GAHQRFGRYG ATQRLCRSSH PRLGSPKFCQ
51 RTRHRSRQQY LYGSFHPQRD WSCPCKIQLG RHHTGSCRAV ADXRDRICER
101 EIRRRQXQCR CLRGKIPSL S IPKYPLKLEQ RYKKNITSS TVPSPNGKNV
151 KLADQRHFKT GVFPDGGKFP NFEKIVKYDT KLDIQELSGG GLPKAKPVFD
201 AKFRWEVURK LNKLTTRQV ERNVQETRRR SQSSQFKAHA QREWENKTGL
251 DFNHFIGGDI NRKGAVTGGH SLTRGVURVI QQTAPDKHG VLSSDDSGN*

35 Further work revealed the complete gonococcal DNA sequence <SEQ ID 463>:

1 TTGGGCATT CCGGCAAAAT ATCCCTTATT CTGTCCATAC TGCGAGTGTG
51 CCGTCCGATG CATGCACACG CCTCAGATT GGcaAACGAT CCCTTTATCC
101 GgCaggttct CGaccGTCAG CATTTCgaac ccgacggcGaa ATACCACTTA
151 TTcggCaGCA GGGGGGAGCT TgcnagcGC aacggccATa tcggattggG
40 201 aaacaTAcAA Agccatcagt tGggccacct gatgattcaa caggcgccgg
251 ttgaaggaaa TAtcgGctac attgtccgct ttccgatca cgggcacaaa
301 ttccattcgc ccttcGAcAA ccaTGcCTCA CATTCCGATT CTGACGAAGC
351 CGGTAGTCCC GTTGAOGGAT TCAGCCTTTA CCGCATCAT TGGGACGGAT
45 401 ACGAACACCA TCCGCGCAG GGCATATGAC GGCCACAGGG CGCGGCTAT
451 CCGGCTCCCA ARGGCGGAG GGAATATAC AGCTACGACA TAAAGGGCGT
501 TGCCCAAAAT ATCGCGCTCA ACCTGACGA CAACCGGAGC ACCGGACAC
551 GGCCTTCCGA CCGTTTCAC AATGCCGGCG CTATGCTGAC GCAAGAGATA
601 GCGCAGCGAT TCAAAOCGCG CACCCGATAC AGCCCGGAGC TGAACAGATC
651 GGGCAATGCC gccGAAGCCT TCAACGCAC TGCAGATATC GTCAAAAACA
701 TCATCGGCGC GGCAGGAGAA ATTGTGCGCG CAGCGGATGAG GTGcagGGT
751 ATAAGCGAAG GCTCAACAT TGCTGTCTAT CACGCGTTGG GTCTGCTTTC
801 CACCGAAAAC AAGATGGGCG GCATCAACGA TTTGGCAGAT ATGGCGCAAC
851 TCAAAAGACTA TGCGCGCAGA GCCATCCGCG ATTGGGCGAT CCAAAACCCC
901 AATGCGGCAC AAGCGATAGA AGCCGTGAGC AATATCTTA TGGCAGCAT
55 951 CCCATCAAA GGGATTGGAG CTGTCCGGGG AAAATACGGC TTGGCGCGCA
1001 TCACGGCACA TCCTGTCAAG CGGTGCGAGA TGGCGCGCAT CGCATTCGCG
1051 AAGGGGAATC CGCGCGTCAG CGACAATTTT GCGGATGCGG CATACGCCAA
1101 AATCCGCTCC CTTTACCAT CCGCAATATC CCGTCAACAT TTGGAGCAGC

5
10
15

```

1151 GTTACGGCAA AGAAAACATC ACCTCCTCAA CCGTGCCGCC GTCAACCGGC
1201 AAAAATGTCA AACTGGCAGA CCAACGCCAC CCGAAGACAG GCGTACCGTT
1251 TGACGGTAAA GGGTTTCGGA ATTTTGAGAA GCACGTGAAA TATGATACGA
1301 AGCTCGATAT TCAAGAATTA TCGGGGGCGC GTATACCTAA GGCTAAGCCT
1351 GTGTTTGATG CGAAACCGAG ATGGGAGGTT GATAGGAAGC TTAATAAATT
1401 GACAACCTCGT GAGCAGGTGG AGAAAAATGT TCAGGAAAGC AGAAGAAGGA
1451 GTCAGAGTAG TCAGTTTAAA GCCCATGCGC AACGAGAATG GGAATAAATA
1501 ACAGGGTAGT ATTTTAAATCA TTTTATAGTT GGTGATATCA ATAGAAGAAGG
1551 CACAGTAACA GAGAGCGATA GTCTAACCCG TGGTGATGTA CGGTGATATC
1601 AACAACTCTC GGCACCTGAT AACATCGGGG TTTTATCAAGC GACAGTGGAA
1651 ATTAATAAAGC CTGATGGAAG TTGGGAGGTT AAAACGAAAA AAGGTGGGAA
1701 AGTGATGACC AAGCACACCA GTTCCCAAAA AGATTGGGAT GAGGCTAGAA
1751 TTAGGGCTGA AGTTACTTCG GCTTGGGAAA GTAGAATAAT GCTTAAGGAT
1801 AATAAATGGC AGGTACAAAG TAAATCGGTT ATTAATAATG AAGGATTTCAC
1851 CGAACCTAAT AGAACAGCAT ATCCCATTTA TGAATAG

```

This corresponds to the amino acid sequence <SEQ ID 464; ORF46ng-1>:

20
25
30

```

1 LGISRKISLI LSLAVCLPM HAHASDLAND PFIRQVLDQR HFEFDGKYHL
51 FGSRGELAXR NGHIGLNIQ SHQLGHLMIQ QAAVEGNIGY IVRFSDHGK
101 FHSFPDNHAS HSDSDSEAGSP VDGFSLYRIH WDGYEHHDPD GYDGPQGGGY
151 PAFKARDIY SYDIKGVQAN IRLNLTNDRS TQRLADRFH NAGAMLTQV
201 GGFKTRATRY PELDRSGNA EAFNGTADI VKNIIAGAGE IVGAGDAVQG
251 ISEGSNIAMV HGLGLSTEN KMARINLAD MQLKDYAAA AIRDWAVQNP
301 NAAQGLEAVS NIFMAAIPIK GIGAVRGKYG LGTTAHPYKA RQMGAIALP
351 KKGSAVSONF ADAAYAKYPS PYHSNRIRSN LEQRYGKANI TSSTVPPSNG
401 KNVKLADQRH PKTGVFDDGC GFNFKEHKVK YDTKLDIQEL SGGGIFPKAK
451 VFDAKPRWEV DRKLNKLTTT EQVEKNVQET RRRSQSSQFK AHAGREWENK
501 TGLDYNHFIG GDINKKGTVT GGHSLTRGDV RVIQQTSAFD KHGVIQATVE
551 IKKPDGSWEV KTKKGGKVM T KHTMFPKDDW EARIRAEVTS AWESRIMLKD
601 NKWQGTSGSG IKIEGFTEPN RTAYPIYE*

```

ORF46ng-1 and ORF46-1 show 94.7% identity in 227 aa overlap:

35
40
45
50
55
60

```

                                10      20      30      40
orf46-1.pep                    AVCLPMHAHASKLANDSFIRQVLDQRHFEFDGKYHLFGSRGELAER
                                |||||
orf46ng-1                      LGISRKISLILSLAVCLPMHAHASDLANDPFIRQVLDQRHFEFDGKYHLFGSRGELAXR
                                10      20      30      40      50      60

                                50      60      70      80      90      100
orf46-1.pep                    QSHIGLGIQSHQLGNLMIQQAIAKGNIGYIVRFSDHGHEVHSFPDNHSHSDSDEAGSP
                                :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf46ng-1                      NGHIGLNIQSHQLGHLMIQQAIAVEGNIGYIVRFSDHGKFHSFPDNHSHSDSDEAGSP
                                70      80      90      100      110      120

                                110     120     130     140     150     160
orf46-1.pep                    VDGFSLYRIHWGDEYHHPADGYDGPQGGGYPAKPGARDIYSYDIKGVQANIRLNLTNDRS
                                |||||
orf46ng-1                      VDGFSLYRIHWGDEYHHPADGYDGPQGGGYPAKPGARDIYSYDIKGVQANIRLNLTNDRS
                                130     140     150     160     170     180

                                170     180     190     200     210     220
orf46-1.pep                    TQRLADRFHNAGSMLTQGVGDFKTRATRYSPELDRLSGNAEAFNGTADIVKNIIGAGE
                                |||||
orf46ng-1                      TQRLADRFHNAGSMLTQGVGDFKTRATRYSPELDRLSGNAEAFNGTADIVKNIIGAGE
                                190     200     210     220     230     240

                                I
orf46-1.pep                    |
                                |
orf46ng-1                      IVGAGDAVQGISSEGSNIAMVHGLGLSTENKMARINLADMAQLKDYAAAAAIRDWAVQNP
                                250     260     270     280     290     300

```

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF46ng-1 shows 87.4% identity over a 486aa overlap with an ORF (ORF46a) from strain A of

N. meningitidis:

10 20 30 40 50 60

	orf46a.pep	LGISRKISLILSILAVCLPMHAHASDLANDSFIQVLDHQHFEPDGKGYHLFGSRGELAER	
	orf46ng-1	LGISRKISLILSILAVCLPMHAHASDLANDPFIQVLDHQHFEPDGKGYHLFGSRGELAXR	
5		10 20 30 40 50 60	
	orf46a.pep	SGHIGLGNISQHLGNLFIQAAIKNGNIGYIVRFSDDHGHVHSPFDNHASHSDSEAGSP	
	orf46ng-1	NGHIGLGNISQHLGNLFIQAAVEGNIGYIVRFSDDHGHVHSPFDNHASHSDSEAGSP	
10		70 80 90 100 110 120	
	orf46a.pep	VDGFSLYRIHWGDEYHHPADGYDGPQGGYPAKPGARDIYSYDIKGVAQNIRNLNLTNRS	
	orf46ng-1	VDGFSLYRIHWGDEYHHPADGYDGPQGGYPAKPGARDIYSYDIKGVAQNIRNLNLTNRS	
15		130 140 150 160 170 180	
	orf46a.pep	TGQRLVDRPHNTGSMLTQGVGDGFKRATRYSPELDNRSGNAEAFNGTADIVKNIIGAAGE	
	orf46ng-1	TGQRLADRFHNGAMLTQGVGDGFKRATRYSPELDNRSGNAEAFNGTADIVKNIIGAAGE	
20		190 200 210 220 230 240	
	orf46a.pep	IVGAGDAVQIGSEGSNIAMVHGLGLLSTENKMARINDLADMAQLKDYAAAAIRDAWQNP	
	orf46ng-1	IVGAGDAVQIGSEGSNIAMVHGLGLLSTENKMARINDLADMAQLKDYAAAAIRDAWQNP	
25		250 260 270 280 290 300	
	orf46a.pep	NAAQGEAVSNIFTAVIPVKIGAVRGYGLGGITAHVPKRSQMGATALPKGKSAVDNF	
	orf46ng-1	NAAQGEAVSNIFTAAIPKIGAVRGYGLGGITAHVPKRSQMGATALPKGKSAVDNF	
30		310 320 330 340 350 360	
	orf46a.pep	ADAAYAKYSPYHSRNRISNLEQRYGKENTISSTVPPSNGKNVKLNKHKHETKVPFDK	
	orf46ng-1	ADAAYAKYSPYHSRNRISNLEQRYGKENTISSTVPPSNGKNVKLADQKHETKVPFDK	
35		370 380 390 400 410 420	
	orf46a.pep	GFPNFEKVDKYDTRINTAVPQVN---PIDEPVFN--PKGSVGSASHWSITARIQYAKLP	
	orf46ng-1	GFPNFEKVDKYDTRINTAVPQVN---PIDEPVFN--PKGSVGSASHWSITARIQYAKLP	
40		430 440 450 460 470	
	orf46a.pep	RQGRIRYIPPKNYSAPSAPLPGPNNGYLDKEFGNEWTGKPSRTKGGEFWDVQLSKTRGQ	
	orf46ng-1	QETRRRSQSSQFKAHQREWEKNTGLDNHFTGGDINKKGTVTGGHSLTRGSDVRVIQOTS	
45		480 490 500 510 520 530	
	orf46a.pep	QETRRRSQSSQFKAHQREWEKNTGLDNHFTGGDINKKGTVTGGHSLTRGSDVRVIQOTS	
	orf46ng-1	QETRRRSQSSQFKAHQREWEKNTGLDNHFTGGDINKKGTVTGGHSLTRGSDVRVIQOTS	

The complete length ORF46a DNA sequence <SEQ ID 465> is:

55	1	TTGGGCATT	CCGCAAAAT	ATCCCTTATT	CTGTCCATAC	TGGCAGTGTG
	51	CCTGCCGATG	CATGCACACG	CCTCAGATT	GGCAACAGAT	TCTTTTATCC
	101	GGCAGGTTCT	CGACCGTCAG	CATTTCGAAC	CCGACGGGAA	ATACCACTTA
	151	TTCCGCGACG	GGGGGGAATC	TGCCGAGCGC	AGCGTCATA	TGGGATTGGG
	201	AAACATACAA	AGCCATCAGT	TGGGCAACCT	GTTTCATCAG	CAGGCGGCGA
	251	TAAAGGAAA	TATCGGCTAC	ATTGTCCGCT	TTTCGATCA	CGGGACGAAA
60	301	GTCATTCCG	CCTTCGACAA	CCATGCTTCA	CATTGCGATT	CTGATGAGC
	351	CGTACTCCG	GTTGAGGAT	TACGCTTTA	CCGATCCAT	TGGACGAGT
	401	ACGACACCA	TCCGCGCAG	GCTATGAGC	GGCGACGGG	CGGGCGTAT
	451	CCGCTCCCA	AAGCGCGAG	GGATATATAC	AGTACGACA	TAAAGAGGCT
	501	TGCCCAAAAT	ATCGCCCTCA	ACCTGACGGA	CAACCGCAGC	ACCGGACAA
65	551	GGCTTGTGCA	CCGTTTCCAC	AATACCGGTA	GATGCTGAC	GCAAGGAGTA
	601	GGCGACGGAT	TCAAAACGCG	CACCCGATAC	AGCCCGGAGC	TGGACGATC
	651	GGGCAATGCC	CCGGAAGCTT	TCAACGCGAC	TGCAGATATC	GTCAAAACAA
	701	TGATCGGCGC	GGCAGGAGAA	ATTGTCCGCG	CAGGCGATGC	CGTGCAGGCT
	751	ATAAGCGAAG	GCTCAACAT	TGCTGTTATG	CACGCTTGG	GTCTGCTTC
70	801	CACCGAAAAC	AAGATGGCGC	GCATCACAGA	TTTGGCAGAT	ATGGCGCAAC

851	TCAAAGACTA	TGCCGCGAGCA	GCCATCCGGC	ATTGGGCGAT	CCTAAACCCG
901	AATGCCGCAC	AAGGCATAGA	AGCGCTCAGC	AATATCTTTA	CGGCATCAT
951	CCCCGTCAA	GGGATTGGAG	CTGTCGGGG	AAATACGGC	TGGGCGGCA
1001	TCACGGCACA	TCCTGTCAAG	CGGTCCGACA	TGGGCGAGAT	CGCATTCGG
1051	AAAGGGAAAT	CGCGCGTCAG	CGACAATTTT	GCCGATGCGG	CATACGCCAA
1101	ATACCCGTCC	CCTTACCATT	CCGCAATAT	CCGTTCAAC	TTGAGCAGC
1151	GTTCGSCAA	AGAAACATC	ACCTCCTCAA	CCGTGCGGCC	GTCAAACGGA
1201	AAGAATGTGA	AACTGGCAAA	CAACGCCAC	CCGAGACCA	AAGTGCCTT
1251	TGACGGTAAA	GGGTTTCCGA	ATTTTGAATA	AGACGTAAAA	TACGATGGA
1301	GAATTATATC	CGCTGACCA	CAAGTGAATC	CTATGATGGA	ACCGCTCTT
1351	AATCCTAAAG	GTTCCTCGG	ATCGGTCAT	TCTTGTCTTA	TAACCTCCAC
1401	AATTCAATAC	GCAAAATTAC	CAAGGCAAGG	TAGATACGA	TATATCCAC
1451	CTAAAAATTA	CTCTCTCTCA	GCACCGCTAC	CAAAAGGACC	TAATATGGA
1501	TATTTGGATA	AATTTGGTAA	TGAATGGACT	AAAGTCCAT	CAAGAACTAA
1551	AGGTCAAGAA	TTTGAATGGG	ATGTTCAATT	GTCTAAACA	GGAAAGAGC
1601	AACTTGGATG	GGCTAGTAGG	GATGGTAAGC	ATTAAATAT	ATCAATTGAT
1651	GGAAAGATTA	CACACAAATG	A		

This corresponds to the amino acid sequence <SEQ ID 466>:

1	LGISRKISLI	LSILAVCLEM	HAHASDLAND	SPIRQVLDRO	HFEPDGKYHL
51	FGSGELAE	SGHIGLNIQ	SHQLGNLFQ	QAAIKNGIY	IURFSDHGE
101	VHSFPDNHAS	HSDSDGASG	VDGFSLYRIH	WDGYEHHPAD	GYDPPGGGY
151	FAFKGARDIY	SYDIKVAQN	IRNLNTNRS	TGRLVDRH	NTGSLTGGV
201	GCDFKATRV	SELDKVAQN	AEAFNGADI	VKNIGRAGE	IVGAGDAVG
251	ISEGNNIAVM	HGLGLLSTEN	KMARINLAD	MAQLKYDAA	AIRDWAQNP
301	NAAQIGIEAVS	NIFTAVIPVK	GIGAVGRKYG	LGGITAHVPK	RSQMGIAPL
351	KGKSAVSDNF	ADAAAYKYP	FYHSNRNRS	LEQRYKENI	TSSTVPPSNG
401	KNVKLANKRH	PKTKVFPDGK	GFNFEDKVK	YDTRINTAVP	QVNPIDEPVF
451	NPKSGVSGAH	SWSITARIQY	AKLPRQGRIR	YIPPKNYSES	APLEKPNNG
501	YLDKFGNEW	KGFSRTKGQE	FEWDVQLSKT	GREQLGWASR	DGKHLNISID
551	GKITHK*				

Based on this analysis, including the presence of a RGD sequence in the gonococcal protein, typical of adhesins, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 56

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 467>:

1	ATGAATATTC	ACACCCCTGCT	CTCCAAACAA	TGGACGCTGC	CGCCATTCCT
51	GCGCAAAACGG	CTGCTGCTGT	CCCTGCTGAT	ACTGCTTGCC	CCCAATGCGG
101	TGTTTTGGGT	TTTGGCAGTG	CTGACGCCCA	CGCCCGGCC	GATTGTCAAT
151	TTGGACTATC	TTCCCGCGCG	GCTGCTGATC	GCCTGCCTT	GCGGTTTCGT
201	CAAAATTGCC	GGCGATTATG	CGTTTGGCT	GGCGGTTTGG	TTTGACGGGC
251	TGATGATGGT	GATCCAACTC	TTCCTTTTAA	TGGATCTCAT	CGCGGCCATC
301	AACCTCGTCC	CCTTCATCCT	GACCGCCCC	GCCCTTATC	AGATAATGAC
351	CGGGCTG...				

This corresponds to the amino acid sequence <SEQ ID 468; ORF48>:

1	MNIHTLLSKQ	WLPFFFLPKR	LLLSLLILLA	PNNAVFWLAL	LTATARPVIV
51	LDYLPALLI	ALPWFREKIA	GVLAFLAVL	FDGLMVIQL	FFPMDLIGAI
101	NLPVFILTAP	APYQIMTGL...			

Further work revealed the complete nucleotide sequence <SEQ ID 469>:

1	ATGAATATTC	ACACCCCTGCT	CTCCAAACAA	TGGACGCTGC	CGCCATTCCT
51	GCGCAAAACGG	CTGCTGCTGT	CCCTGCTGAT	ACTGCTTGCC	CCCAATGCGG
101	TGTTTTGGGT	TTTGGCAGTG	CTGACGCCCA	CGCCCGGCC	GATTGTCAAT
151	TTGGACTATC	TTCCCGCGCG	GCTGCTGATC	GCCTGCCTT	GCGGTTTCGT
201	CAAAATTGCC	GGCGATTATG	CGTTTGGCT	GGCGGTTTGG	TTTGACGGGC
251	TGATGATGGT	GATCCAACTC	TTCCTTTTAA	TGGATCTCAT	CGCGGCCATC
301	AACCTCGTCC	CCTTCATCCT	GACCGCCCC	GCCCTTATC	AGATAATGAC
351	CGGGCTGTTG	CTGCTGTATA	TGCTGCGGAT	GCGGTTTGTG	TTGCAGAAAG

25

30

35

40

45

50

55

60

501 GATGGCCAAT ATCTTCGGGG CAAACAACTT CTATTACGCC AAAAGTCAGG
 551 CGATGCTCTA CACCGTCAGC CAGAAATGCC ACTTTATTAC CGCCGCGCTG
 601 GTCGATCCCG TCTTCTCTCC CTGGGCAAT CAACAGCGTG CGCCACGCA
 651 TCTGAACGAG CGAAATCTC AAAAAATCCT CTTTATCGTC GCGCAATCTT
 701 GGGGCTGCC GGCCAATCCC GAACCTCAA ACGCCACTT TGCCAACTG
 751 CTGGCGCAAA AAGANCGTTT TCGGTTTGG GAAGCGCGCA GTTTTCCCTT
 801 CATCGCGCGG ACGATCGAAG CGGAATATGC CGAATGTGT GCTCATCGGG
 851 GTTTGCGCGG GTTCGCACTG CGCCGCGCGC CGACGAAAA ATTTGCGCGC
 901 TGCCTCCCA ACCGTTTGAA ACAGAGAGT TACGCACCT TTGCGATGCA
 951 CGCGCGCGCG ACTTCGCTTT ACAGACGCTT CAGCTGGTAT CGAGGCGCGG
 1001 GCTTTCAGAA AATCAAACC CGCGAAACC TGATCGGTAA AAAACCTGCG
 1051 GCCATTTTCG GGGCGGTGTG CGACAGCGAG CTGTTGCGCG AAGTGTGCGG
 1101 ANTTTTCAA AAACACGACA AGGGACTGTT TTAATGGATG ACGCTGACCA
 1151 GCCACGCGGA CTATCCCGAA TCNGACATT TCAACACAG GCTCAATGC
 1201 ACCGAATATG GCTGCGCCGC CGAAACGAGC NTCTGCGCGA ATTTACGCTC
 1251 GCACACCCAA TTCTTCGACC AACTGGCGGA TTTGATCAA CGCCCGGAAA
 1301 TGAAGGCAC GGAAGTATC ATGCTGCGCG ACCATCCGCG GCCCGTGGC
 1351 AACCTCAATG AAACCTTCG CTACCTCAA CAGGGGACG TCGNCTGCT
 1401 GAACCTCAA ATCAATAA

20 This encodes a protein having amino acid sequence <SEQ ID 472>:

1 MNIHTLLSKQWTLPPFLPKR LLSLLILLX PNAVFWVIAL LTATAPFVN
 51 LXYLPAALLI ALPWRXKIX GVLAAXLAVL FDGLMMVQL PFMDLIGAN
 101 NLVPEIXTAP ALYQIMTGLL LLYMLAMEFV LKAAAKTDF RHIAACAADV
 151 VAAGYFTGHL SKYDRGRMAN IFGANNFYA KSQAMLYTVS QNADFITAGL
 201 VDPVFLPLGN QQRATHLNE FKSQKILFV AESWGLPAP ELQNATFAKL
 251 LAQKXRFVSW ESGSFFFIGA TIEGEMRELK AYGLRGFAL RRAPEKFAK
 301 CLPNRLKQEG YATFAMHAGG SLYDRFSWY PRAGFQEKTK AENLIGKKT
 351 AIFGVCDS E LFGVSAXFK KHKDGLFYWM TLTSHADYFE SDIFNHLK
 401 TEYGLPAETD XCRNFSLHTQ FFDQLADLIQ RPEMKGTVEI IVGDHPEFPVG
 451 NLNETFRYLK QGHVXWLNFK IK*

ORF48a and ORF48-1 show 96.8% identity in 472 aa overlap:

		10	20	30	40	50	60
orf48a.pep	MNIHTLLSKQWTLPPFLPKRLLSLLILLX PNAVFWVIAL LTATAPFVN						
orf48-1	MNIHTLLSKQWTLPPFLPKRLLSLLILLX PNAVFWVIAL LTATAPFVN						
		10	20	30	40	50	60
orf48a.pep	ALPWRXKIXGVLAAXLAVLFDGLMMVQLFPFMDLIGAN						
orf48-1	ALPWRXKIXGVLAAXLAVLFDGLMMVQLFPFMDLIGAN						
		70	80	90	100	110	120
orf48a.pep	LYLPEIXTAPALYQIMTGLL LLYMLAMEFV LKAAAKTDF RHIAACAADV						
orf48-1	LYLPEIXTAPALYQIMTGLL LLYMLAMEFV LKAAAKTDF RHIAACAADV						
		130	140	150	160	170	180
orf48a.pep	VAAGYFTGHL SKYDRGRMAN IFGANNFYA KSQAMLYTVS QNADFITAGL						
orf48-1	VAAGYFTGHL SKYDRGRMAN IFGANNFYA KSQAMLYTVS QNADFITAGL						
		190	200	210	220	230	240
orf48a.pep	VDPVFLPLGN QQRATHLNE FKSQKILFV AESWGLPAP ELQNATFAKL						
orf48-1	VDPVFLPLGN QQRATHLNE FKSQKILFV AESWGLPAP ELQNATFAKL						
		250	260	270	280	290	300
orf48a.pep	LAQKXRFVSW ESGSFFFIGA TIEGEMRELK AYGLRGFAL RRAPEKFAK						
orf48-1	LAQKXRFVSW ESGSFFFIGA TIEGEMRELK AYGLRGFAL RRAPEKFAK						
		310	320	330	340	350	360
orf48a.pep	CLPNRLKQEG YATFAMHAGG SLYDRFSWY PRAGFQEKTK AENLIGKKT						
orf48-1	CLPNRLKQEG YATFAMHAGG SLYDRFSWY PRAGFQEKTK AENLIGKKT						
		370	380	390	400	410	420
orf48a.pep	AIFGVCDS E LFGVSAXFK KHKDGLFYWM TLTSHADYFE SDIFNHLK						
orf48-1	AIFGVCDS E LFGVSAXFK KHKDGLFYWM TLTSHADYFE SDIFNHLK						
		430	440	450	460	470	480
orf48a.pep	TEYGLPAETD XCRNFSLHTQ FFDQLADLIQ RPEMKGTVEI IVGDHPEFPVG						
orf48-1	TEYGLPAETD XCRNFSLHTQ FFDQLADLIQ RPEMKGTVEI IVGDHPEFPVG						
		490	500	510	520	530	540
orf48a.pep	NLNETFRYLK QGHVXWLNFK IK*						
orf48-1	NLNETFRYLK QGHVXWLNFK IK*						

		370	380	390	400	410	420
orf48a.pep		LFGEVSAXFKKHKGLFYWMTLTSHADYPESDIFNHRKCTEYGLPAETDCXCRNFSLHTQ					
5	orf48-1	LFGEVSAXFKKHKGLFYWMTLTSHADYPESDIFNHRKCTEYGLPAETDCXCRNFSLHTQ					
		370	380	390	400	410	420
		430	440	450	460	470	
orf48a.pep		FFDQLADLIQRPEMKGTEVIVGDHPPVGVNLETFRYLKQGHVXWLNFKIKX					
10	orf48-1	FFDQLADLIQRPEMKGTEVIVGDHPPVGVNLETFRYLKQGHVXWLNFKIKX					
		430	440	450	460	470	

Homology with a predicted ORF from *N. gonorrhoeae*

ORF48 shows 97.5% identity over a 119aa overlap with a predicted ORF (ORF48ng) from *N.*

15 *gonorrhoeae*:

	orf48.pep	MNIHTLLSKQWTLPPFLPKRLLSLILLAPNAVFWVLALLTATARPVINDYLPAAALI	60
	orf48ng	MNIHALLSEQWTLPPFLPKRLLSLILLAPNAVFWVLALLTATARPVINDYLPAAALI	60
20	orf48.pep	ALPWRVVKIAGVLAFLAVLFDGLMMVQLFPFMDLIGAINLVPILITAPAPYQIMTGL	119
	orf48ng	ALPWRVVKIAGVLAFLAVLFDGLMMVQLFPFMDLIGAINLVPILITAPAPYQIMTGL	120

The ORF48ng nucleotide sequence <SEQ ID 473> was predicted to encode a protein having amino acid sequence <SEQ ID 474>:

1	MNIHALLSEOWTLPPFLPKRLLSLILLAPNAVFWVLALITATARPVNL
51	LDVLPALILI ALPWRVVKIA GVLAFLAVL FDGLMMVQL FPMDLIGAI
101	NLVPILITAP APYQIMTGLL LLYLAMPFV LQAAVKTFD RHIAVCAAVV
151	AAARYFTGPF ELLRTGGRWV YVQHRLLLS GSRASFRRQ KADVLRLGN
201	PYASMGNGG..

30 Further work identified the complete gonococcal DNA sequence <SEQ ID 475>:

1	ATGAATATTC	ACGCCCTGCT	CTCCGAACRA	TGACGCTGCG	CGCATTCCT
51	GCGGAACGG	CTGCTGCTGT	CCCTGCTGAT	ACTGCTGCC	CGCATTCCT
101	TGTTTGGGT	TTTGCCACTG	CTGACGCCA	CGCGCCGCC	GATTGTCAAT
151	TGGAATACC	TTCCCGCCG	GCTGCTGATC	GCCTCGCTT	CGGTTTCTT
201	CAAAATGCG	GGCGTATTGG	CGTTTGGCC	GGCGSTTTTG	TTTACCGGGC
251	TGATGATGT	GATCCAATC	TTCCCTTTTA	TGACCTCAT	CGCGCCATC
301	AACCTCGTC	CCTTCATCT	GACCGCCCG	GCCTCTTATC	AGATAATGAC
351	CGGCTGTGT	CTGCTGTATA	TGCTGGCAT	CGGTTTGTG	TTGCAAAAG
401	CGCGCGTCA	AAACGACTTC	CGACACATTG	CGCTCTGTGC	CGCGCTGTG
451	CGCGCAGCG	GCTATTTCAC	CGGCACTTTG	AGTTACTACG	ACCGGGGGCG
501	GATGCGCAAT	ATCTTCGGCG	CAACACACTT	CTATTACGCC	aAAAGTCAGG
551	CGATGCTCTA	CACGCTCAGC	CAGAATGCG	ACTTTATTAC	CGCGGcctg
601	GTGACCCCG	TCTTCCTCCC	CTTGGGCAAT	CAGCAGCGTG	CGCGCACGG
651	GTCAGATAG	CGAATATCT	AAATAATCT	CTTATCGTIC	CGGCAATCTT
701	CGGCTGCGC	GGCGATATCC	GACCTTCARA	ACGCCACTT	TGCAACAATG
751	CTGCGCAAA	AAGACCGTTT	TTGCGTTTGG	GAAGCGGCA	GTTTCCCTT
801	CATCGCGCG	ACGCTCGAAG	GCGAAATGCG	CGAATTGTGC	GCCTACGGCG
851	GTTCGCGCG	GTTCCGACTG	CGCGCGCGCG	CGACGAAAA	ATTGCCCCG
901	TGCTCCCCA	ACCGTTTGAA	ACGAAGAGT	TACGCAACCT	TTGCGATGCA
951	CGCGCGGGT	AGTTGCTGTT	ACGACCGGTT	CAGCTGGTAT	CGAGGGGGG
1001	GCTTTCAAAA	AATCAAAACC	GCGCAAAACC	TGATCGGTAT	AAAAACCTCG
1051	GCCATTTTCG	GCGCGGTGTG	CGACAGOGAG	CTGTTGCGCG	AAGTGTGCGG
1101	ATTTTTCAAA	AAACACGACA	AGGACCTGTT	TACTTGATAG	ACGCTGACCA
1151	GCCACGCGGA	CTATCCCGAA	TCGACATTIT	TCAACACAG	GCTCAAATGC
1201	ACCGAATACG	GCCTGCGCGC	CGAAACCGAC	CTCTGCGCGA	ATTTACAGCT
1251	GCACACCGAA	TcttcgACC	AACCTGGCGA	TTTGATCGGA	CGCGCCGAAA
1301	TGAAGCGAC	CGAATATCT	ATGCTGCGCG	ACCATCGCGC	CGCGCTCGCG
1351	AACTCAATG	AAACCTTCG	CTACTCATA	CAGGACACG	TGCGCTGCGT
1401	GCACCTCAA	ATCAATAA			

60 This encodes a protein having amino acid sequence <SEQ ID 476; ORF48ng-1>:

	1	MNIHALLSEQ	WTLPPFLPKR	LLLSLLILLA	FNADVFWLAL	LTATARPVFN
	51	LDYLPALLT	ALPWRFKIA	GVLAFWPAVL	FDGLMMVIO	FFPMDLIGAI
	101	NLVPFILTAP	APYQIMTGLL	LLYMLAMPFV	LKQAAVKTDF	RHIAVCAAVV
5	151	AAAGYFTGHL	SYVDGRGMAN	IFGANNFYFA	KSQAMLYTVS	QNADFITAGL
	201	VDPVFLPLGN	QCRATRLSE	PKSQKILFIV	AESWGLGPN	ELQNATFAKL
	251	LAQKDRFSVM	ESGSPFFIGA	TVEGEMRELC	AYGGLRGFAL	RRAPDEKFA
	301	CLPNRLKQEG	YATFAMHGAG	SSLYDRFSWY	PRAGFQKIKT	AENLIGKTC
	351	ALFGGVCDSE	LPGEVSAPFK	KHDKGLFYWM	TLTSHADYFE	SDIFNHLKAC
10	401	TEYGLPAETD	LCRNFSLHTQ	FFDQLADLIR	RPENMGTEVI	IVGDHPPFVG
	451	NLNETFRYLK	QGHVWLHFK	IK*		

ORG48ng-1 and ORF48-1 show 97.9% identity in 472 aa overlap:

		10	20	30	40	50	60
	orf48-1.pep	MNIHTLSKQWTLPPFLPKRLLLSLLILLAPNAVFWLALLTATARPVFNLDYLPALLI					
15	orf48ng-1	MNIHALLSEQWTLPPFLPKRLLLSLLILLAPNAVFWLALLTATARPVFNLDYLPALLI					
		10	20	30	40	50	60
		70	80	90	100	110	120
20	orf48-1.pep	ALPWRFKIAGVLAFLAVLFDGLMMVIOQLFPFMDLIGAINLVPFILTAPAPYQIMTGLL					
	orf48ng-1	ALPWRFKIAGVLAFLAVLFDGLMMVIOQLFPFMDLIGAINLVPFILTAPAPYQIMTGLL					
		70	80	90	100	110	120
		130	140	150	160	170	180
25	orf48-1.pep	LLYMLAMPFVLQKAAAKTDFRHIACAAVVAAGYFTGHLSSYDGRGMANIFGANNFYFA					
	orf48ng-1	LLYMLAMPFVLQKAAVKTDFRHIACAAVVAAGYFTGHLSSYDGRGMANIFGANNFYFA					
		130	140	150	160	170	180
30		190	200	210	220	230	240
	orf48-1.pep	KSQAMLYTVSQNADFITAGLVDPVFLPLGNQCRATRLNEPKSQKILFIVAESWGLPANG					
	orf48ng-1	KSQAMLYTVSQNADFITAGLVDPVFLPLGNQCRATRLSEPKSQKILFIVAESWGLPANG					
		190	200	210	220	230	240
35		250	260	270	280	290	300
	orf48-1.pep	ELQNATFAKLLAQKDRFSWESGSPFFIGATVEGEMRELCAYGGLRGFALRRAPDEKFA					
	orf48ng-1	ELQNATFAKLLAQKDRFSWESGSPFFIGATVEGEMRELCAYGGLRGFALRRAPDEKFA					
40		250	260	270	280	290	300
		310	320	330	340	350	360
	orf48-1.pep	GLFNRLKQEGYATFAMHGAGSSLYDRFSWYPRAGFQKIKTAENLIGKKTGAIFGGVCDSE					
45	orf48ng-1	CLFNRLKQEGYATFAMHGAGSSLYDRFSWYPRAGFQKIKTAENLIGKKTGAIFGGVCDSE					
		310	320	330	340	350	360
		370	380	390	400	410	420
50	orf48-1.pep	LPGEVSAPFKKHDKGLFYWMTLTSHADYFESDIFNHLKCTEYGLPAETDLCRNFSLHTQ					
	orf48ng-1	LPGEVSAPFKKHDKGLFYWMTLTSHADYFESDIFNHLKCTEYGLPAETDLCRNFSLHTQ					
		370	380	390	400	410	420
55		430	440	450	460	470	
	orf48-1.pep	FFDQLADLIRPEMKGTEVIIVGDHPPFVGNLNETFRYLKQGHVWALHFKIKX					
	orf48ng-1	FFDQLADLIRPEMKGTEVIIVGDHPPFVGNLNETFRYLKQGHVWALHFKIKX					
		430	440	450	460	470	

Based on this analysis, including the presence of a putative leader sequence (double-underlined) and two putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N. meningitidis* and *N. gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 57

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 477>:

```

1  .GTGAGCGGAC GTTACC CGCG TTTGGATCGC GTTTCACAAA TCATCATCGT
51 TACTTTGAGT ATCGCCACGC TTGCCGCGCG CGGCATCGCT ATGTGCGCGG
101 GTATTGCAGAT GCAGTCCGAT TTTATCGAGC CGACACCGTG GACGCTTGCC
151 GGTTTGGGCTT TCCTGATCGC GCTGATGGCG TGGATGCCCG CGCGATTGAC
201 AATTTTCGCGC ATCAATTCTT TGTGGGTAA CCAAAAACAA CGCATCAATC
251 CTTCCGAATA CGCGACGCGG ATTTTGAAT TCAACGTCGG TTATATCGCC
301 AGTGCGGTTT TGGCTTTGGT TTTCTTGCA CTGGCGCG .G TAGCGCGGAA
351 CGGCAACGGC GA .ACATGTC AGATGGCGGG CGGCAATAT ATACGGCGAAT
401 TGATCAATAT GTACGCC .

```

This corresponds to the amino acid sequence <SEQ ID 478; ORF53>:

```

1  .VSGRYRALDR VKIILVITLS IATLAAAGIA MSRGMMQSD FIEPTFWTLA
51 GLGFALALMG WMPAPIEISA INSLWVTEQK RINPSEYRDG IFEFNVGYIA
151 SAVLALVELA LGXVAPNNGN XTVQMAGGKY GQLNMHYA .

```

Further work revealed the complete nucleotide sequence <SEQ ID 479>:

```

1  ATGTCCGAAC AACATATTTT GACTTGGAAA AGTAAATCA ACGCATTGGG
51 TCCGGGGATC ATGATGGCTT CGCGCGCGGT CGCGGGTTTG CACCTGATTG
101 CCTCGACGCA GCGGGGCGCG CTTTACGGCTT GGCAGATCGC GCTCATCATC
151 ATCCTGACCA ACCTCTTCAA ATACCGCTTT TTCCGCTTCA CGCGCGATTA
201 CACGCTGGAC ACGGGCAAGA GCCTGATTGA AGGTTATGCC GAGAAAAGCC
251 GCGTTTATTT GTGGGTATTC CTGATTTTGT GCATCTCTCT CGCGACGATT
301 AAGCGGGGCG CGSTCGCCAT TGTAAACCGC GCCATCGTCA AAATGGCGAT
351 TCCTCGCTGT ATGTTTGATG CGCGCACGCT TGCGGCTGT ATTATGCGAT
401 CTTGCGCTGAT TATTTTGGTG ACGGACGCTT ACCGCGCTTT GGATCGCGTT
451 TCCTAAATCA TCATCGTTAC TTTGAGTATC GCCACGCTTG CGCGCGCGG
501 CATCGCTATG TCGCGCGGTA TCGAGATGCA GTCCGATTTT ATCGAGCCGA
551 CACCGTGGAC GCTTGCCTGT TTGGGCTTCC TGATCGCGCT GATGGGCTGG
601 ATGCGCGCGC GATTGAAAT TTCCGCGCAT AATTCTTTGT GGGTAAACGA
651 AAAACACACG ATCAATCCTT CGAATACCG CGACGGGATT TTTGATTCCA
701 AGTCTGGTTA TATCGCCAGT GCGGTTTGG CTTTGGTTT CCTTGCACTG
751 GCGCGCTTTG TGCAATACGG CAACGGCGAA CGAGTGCAGA TGGCGGGCGG
801 CAAATATATC GGGCAATTGA TCAATATGTA CGCGGTTACC ATCGCGGCT
851 GGTGCGCGCC GCTGCTGGCG TTTATCGGCT TTGCTGTAT GTACGGCACG
901 ACGATTACCG TCGTGGACGG CTATGCGCGT GCCATTGCC AACCGGTGGG
951 CCTGCTCGCG GSAAGACAA AACGGGGGCA CGCGGAATTC TTTGCTGGGA
1001 ATATTGUGGT GCGGGGACG GGTTTGGCGG TGATTTTCTG GTTGACGATC
1051 GTAATCGCGA ATCTGCTCAA ATTTGCGATG ATTGCGGCT TTTGCTCGCG
1101 CCTGTGTTT GCTCGGCTGA ATTACGCTTT GTTAAAGGT GATGAAACAC
1151 ACAAACTCAC ATCAGGTATG AATGCCCTTG CATTTGCAGG CTTGATTATT
1201 CTGACCGGTT TTACCGTTTT GTTCTTATTG AATTGCGCGG GAATGTTCAA
1251 ATGA

```

This corresponds to the amino acid sequence <SEQ ID 480; ORF53-1>:

```

1  MSEQHISTWK SKINALGPGI MMASAAVGGG HLIASTQAGA LYGMQIALII
451 ILTNLFKYPP FRFSAHYTLD TKGSLIEGYA EKSRYLVWF LILCILSATI
101 NAGAVAIYTA AIVKMAIPLS MFDAGTVAAL IMASCLIIIV SGRYALDRV
151 SKIILVITLSI ATLAAAGIAM SRGMQMGSDF IEPTFWTLAG LGLIALMGW
201 MPAPIEISAI NSLWVTEQR INPSEYRDI FEFNVGYIAS AVLALVELAL
501 GAFVQYNGEG AVQMGAGKYI GQLNMIAVT ICGWSRPLVA FIAFACMYGT
301 TTTVUCYAR ALAEVRLLR GKDKTGNAEF FAWNILWZSG GLAVIEFDG
351 VMANLKKFAM IAAFVSAPVF AQLNYRLVKG DEKHKLTSGM NALALAGLIY
401 LTGETVLELL NLAGMFK*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

55 ORF53 shows 93.5% identity over a 139aa overlap with an ORF (ORF53a) from strain A of *N.meningitidis*:

	1	ATGCGCGGAC	ACATCATTTT	CAGCTGGAAA	AGTAAAGTA	ACGCAATGGG
	51	ACCGCGGATG	ATGATGGCTT	GCGGCGGCTT	CGGCGGTTGC	CCAGCAGTTG
25	101	CTCGGACGCA	GGCGGCGCGG	CTTACCGGCT	GGCAGATCGC	GCTCATCATC
	151	ATCCTCGACG	ACCTCTTCAA	ATTACCGGCT	TTTCGCTGTA	CGCGCGATTG
	201	CACGCTGGAC	ACGCGCAGGA	CGCATGTGGA	AGGTTATGCC	GAGAAAGATG
	251	GGCTTTATTT	GTGGGTTATC	CTGATTTGTT	GCATCTCTGC	GGCAGAGTTT
	301	AAACGCGGCG	CGTGCGCAT	TGTAAACGCC	CGCATCGTCA	AAATGGCGAT
	351	TCCTCGCTGT	GTGTTTGATG	GGCGCAGCGT	TGCCTGCTTG	ATTATGGCAT
30	401	CTCGCTCGAT	TTGTTTGGTG	ACGCGCGGCT	ACCGCGCTTT	GGATCGCGTT
	451	TCGCAAAAGT	CGGCGCTTAC	TTTGAGTATG	CGGCGCTGTT	CGGCGCGGTT
	501	CATCGTATG	TGCGCGGGTA	TCGACGAGTA	GTCCGATGTT	ATGACGCGCT
	551	CATCGCTGGC	CTGTCGCGST	TGGGCTGCTG	TGATCGCGGT	GATGGCTGGT
35	601	ACGCGCGGCG	CGATTGGAAT	TGCCGCATCT	AAATTCTTGT	GGGTAAACGA
	651	AAACACACGC	ATCACTCACT	CGGATACCG	CGACGGGATT	TTTGAATTCA
	701	AGTCGGGTGA	TGTCGCCAGT	GGGCTTTTGG	CTTGCACTGT	CTTGCACTGT
	751	GGCGCGTTTG	TGCATTAACG	CAACGGGAGA	CGAGTCGAGA	TGGCGGCGGG
	801	CAATATATGT	GGGCAATTGA	TCAATATGTA	CTGCGTTAAC	ATCGCGCGCT
40	851	GTCGCGCCCG	CTGCTGTGCG	TTATCGCGTT	TTGCTGTGTA	GTACGCGCGT
	901	ACGATTACCG	TTGTGGACGG	CTTACGCGTG	GCATATCGCG	AACCGTCGCG
	951	CTATCTGCGG	GGAAAGACGA	AAACGGGACA	CGCGGAATTT	TTTGCTGCGA
	1001	CTTCTTGGGT	GGCGGCGAGC	GGTTTGGGAG	GGATTCTCTG	TTGTGACGGA
	1051	GTATGACGCA	ATCTGCTGAT	ATTCGATGAT	GGTGGCGGCT	TTGTGCTGCT
	1101	CCCTGCTGTA	ATCTGCTGAT	ATTACCGGTT	CGTAAAGGAT	ATGAAAGGAT
45	1151	ACAAACTCAC	ATCAGTATGT	AAATCGCCTT	GTGTCGAGG	CTTGATTTAT
	1201	CTGACCGGTT	TTACCGGTTT	GTCTCTATTG	AAATTGGGCG	GAATGTTCAA
	1251	ATGA				

50	1	ISEQHSITXKX	SNKALNGPGI	TMSAAVVGGS	HLIATSPQCA	LYGQWQIALII
	51	ILNFILKYFP	FRFSAHYTLD	GKASJAEV	EXSRVYLWVF	LIILCSATTI
	101	NAGAWAVTVA	ATVMAIAPSL	MFDGATQATL	IMASCLILVI	SGRYALDRWV
	151	SKIIIVLTSL	ATLAAAGATM	SRGCMGOSDF	IEPTFTWTLG	LGFLIALMGW
	201	MPAFIEISAI	NSLWVTEKQR	INSEYRDEGI	DFPNVNGYIA	AVLAVLFLAL
55	251	GAFVGYNGAE	AVCMAGQYKI	QOLINMTAL	IGGWSRFLVA	FIAFACHYMI
	301	TIVVUDYAR	AIAPVRLLR	GKQTKGTH	FAWNIWGS	GLAVTFWFG
	351	PMANLKEP	IAVPSAVTF	INLVYRWG	DEKIMTGM	NALALGLTY
	401	LTAFIVLFL	NALMGK ⁶			
	451					

```

60      10      20      30      40      50      60
orf53a.pep  MSEQHISTWKSINALGPGIMMASAAVGGSHLIASTQAGALYGWQIALIIILTNLFKYFF
orf53-1     MSEQHISTWKSINALGPGIMMASAAVGGSHLIASTQAGALYGWQIALIIILTNLFKYFF
      10      20      30      40      50      60
65      70      80      90     100     110     120

```

-284-

	orf53a.pep	FRFSAHYTLDTGKSLIEGYAEKSRVYLWVFLILCILSATINAGAVAIVTAIAIVKMAIFSL	
	orf53-1	FRFSAHYTLDTGKSLIEGYAEKSRVYLWVFLILCILSATINAGAVAIVTAIAIVKMAIFSL	
5		70 80 90 100 110 120	
	orf53a.pep	130 140 150 160 170 180	
	orf53-1	130 140 150 160 170 180	
10		190 200 210 220 230 240	
	orf53a.pep	190 200 210 220 230 240	
	orf53-1	190 200 210 220 230 240	
15		250 260 270 280 290 300	
	orf53a.pep	250 260 270 280 290 300	
	orf53-1	250 260 270 280 290 300	
20		310 320 330 340 350 360	
	orf53a.pep	310 320 330 340 350 360	
	orf53-1	310 320 330 340 350 360	
25		370 380 390 400 410	
	orf53a.pep	370 380 390 400 410	
	orf53-1	370 380 390 400 410	
30		420 430 440 450 460	
	orf53a.pep	420 430 440 450 460	
	orf53-1	420 430 440 450 460	
35		470 480 490 500 510	
	orf53a.pep	470 480 490 500 510	
	orf53-1	470 480 490 500 510	

Homology with a predicted ORF from *N.gonorrhoeae*

ORF53 shows 92.1% identity over a 139aa overlap with a predicted ORF (ORF53ng) from *N. gonorrhoeae*:

	orf53.pep	VSGRYRALDRVSKIIIVTSLIATLAAAGIA	30
40	orf53ng	AAIVKMAIPSLMFDAGTVAALIMASCLIIIVSGRYRALDRVSKIIIVTSLIATLAAAGIA	91
	orf53.pep	MSRGQMQSDPIETPTWTLAAGLGLIALMGWMPAPIEISAINSLWVTEKQKRNPFSEYRDG	90
45	orf53ng	MSRGQMQSDPIETPTWTLAAGLGLIALMGWMPAPIEISAINSLWVTEKQKRNPFSEYRDG	151
	orf53.pep	IFEFNVGYIASAVLALVFLALGKVPNGNGKTVQMGAGKYNGQLINMYA	139
50	orf53ng	IFEFNVGYIASAVLALVFLALGAFVQYNGEAVQMGGGKYNGQLINMYAVTTIGGSRPLV	211

An ORF53ng nucleotide sequence <SEQ ID 483> was predicted to encode a protein having amino acid sequence <SEQ ID 484>:

	1	MPKSKCVYLW VFIILCIASA TINAGAVAIV TAAIVKMAIP SLMFDAGTVA
55	51	ALIMASCLII LVSGRYRALD RVSKIIIVTL SIATLAAAGI ANSRGMQMP
	101	DFIEPTPTWL AGLGLIALMGWMPAPIEIS AINSLWVTEK QRINPFSEYRD
	151	GIFDNVGYI ASAVLALVEL ALGAFVQYNG EAVQMGGGK YIGQLINMYA
	201	VTIGGSRPL VAFIAFACMY GAATVVDGY ARAIAEPVRL LRGDRTARP
	251	IVLLEKLGGK HRPGRFLV*

Further analysis revealed further partial DNA gonococcal sequence <SEQ ID 485>:

60	1	..aagaAAGCTG	CGGTTTATTT	GTGGGTTTTT	TTGATTTTGT	GTATCGCCTC
	51	CGCCACGATT	AACGCGGGCG	CGGTCCGCAT	TGTAACGCC	GCCATCGTCA
	101	AAATGGCGAT	TCCCTGCTG	ATGTTTGATG	CCGGCACGGT	TGCCGCTTGT

-285-

5	151	ATTATGGCAT	CCTGCGTGAT	TATTTTGGTG	AGCGGACGTT	ACCGCGCTTT
	201	GGATCGTGT	TCCAAAATCA	TCATTGTTAC	TTTGAGCATC	GCCACGCTTG
	251	CGCGCGCGG	CATCGCTATG	TCGCGCGGTA	TGCAGATGCA	GCCCGATTTT
	301	ATCGAGCGCA	CACCGTGGAC	GCTTGGCGGT	TTGGGCTTCC	TGATCGCGCT
	351	GATGGGCTGG	ATGCCCGCGC	CGATCGAAAT	TTCGCGCATC	AATTCTTTGT
	401	GGTAACCGA	AAACCAACGC	ATCAATCCTT	CTGAATACCG	CGACGGGATT
	451	TTGGAATTCA	ACGTGCGTTA	TATCGCcaat	GCGGTTTTGG	CTTTGGTTTT
	501	CCTTGCACGT	GGCGCGTTTG	TGCATATCGG	CAACGCGGAA	GCAGTGCAGA
	551	TGCGCGCGG	CAAAATATTC	GGCAATTTGA	TTAATATGTA	TGCGGTAAAC
10	601	ATCGCGGCT	GCTTCTGCTC	GCTGGTGGCC	TTATCTGCTT	TGCGGTATAT
	651	GTACGCGCAG	ACGATTACCG	TTGTGGACGG	TTATCGCGGT	GCCATTGCGC
	701	AACCCGTGCG	CCTGCTGCGC	GCGACGGATA	AAACCGCGAA	CGCCGAGTTG
	751	TTTgctTGGa	ATATTTGGGT	GGCGGGCAGC	GTTTGGCGGG	TGATTTTCTG
	801	GTTTGACggc	gcaaTGGCgG	AACTgcTCAA	ATTTCGCGATG	ATtgccgcCT
15	851	TTGTGTCCGG	CCCTGTGTTT	GCTTGGCTCA	ACTACCGCCT	CGTCAAGGGG
	901	GACAAACGCC	ACAGGCTTAC	CGCCGCTATG	AACGCGCTTG	CCATTGTGCG
	951	CCTGCTCTAC	CTGGCCGGGT	TTGCGGTTTT	GTTCTCTGTT	AACCTTACCG
	1001	GACTTTTGGC	ATAG			

This corresponds to the amino acid sequence <SEQ ID 486; ORF53ng-1>:

20	1	..KKSCVYLWVF	LILCIASATI	NAGAVAIUTA	AIVKMAIPSL	MFDAGTVAAL
	51	IMASCLIIILV	SGRYRALDRV	SKIIIVTISI	ATLAAAGIAM	SRGMQMDFDF
	101	TEPTFWLFG	LGFLIALMGW	MEATIEISAI	NSLWTEKQR	INPSETRDGI
	151	FDNNGVYIAS	AVLALVFLAL	GAFVQYNGGE	AVQMAGKGYI	GQLINMYAVT
	201	IGGWSRPLVA	FIAFACMYGT	TIIVVDGYAR	IAAEFVRLLR	GRDKTGNREL
25	251	FAWNIVWAGS	GLAVIFWFDG	AMAECLKFAM	IAAFVSAFPV	AWLNYRLVKG
	301	DKRHRLTAGM	NALAIIVGLLY	LAFVFLFLL	NLTGLLA*	

ORF53ng-1 and ORF53-1 show 94.0% identity in 336 aa overlap:

		60	70	80	90	100	110
orf53-1.pep	ILTNLFKYPFFRFS	SAHYTLDTGKSL	IEGYAEKSRVYL	WVFLILCSAT	INAGAVAIUTA		
orf53ng-1							
					KKSCVYLWVF	LILCIASATI	NAGAVAIUTA
					10	20	30
		120	130	140	150	160	170
orf53-1.pep	AIVKMAIPSL	MFDAGTVAAL	IMASCLIIILV	SGRYRALDRV	SKIIIVTISI	ATLAAAGIAM	
orf53ng-1	AIVKMAIPSL	MFDAGTVAAL	IMASCLIIILV	SGRYRALDRV	SKIIIVTISI	ATLAAAGIAM	
		40	50	60	70	80	90
		180	190	200	210	220	230
orf53-1.pep	SRGMQMDFDF	TEPTFWLFG	LGFLIALMGW	MWPAFIEISAI	NSLWTEKQR	INPSETRDGI	
orf53ng-1	SRGMQMDFDF	TEPTFWLFG	LGFLIALMGW	MWPAFIEISAI	NSLWTEKQR	INPSETRDGI	
		100	110	120	130	140	150
		240	250	260	270	280	290
orf53-1.pep	FDNNGVYIAS	AVLALVFLAL	GAFVQYNGGE	AVQMAGKGYI	GQLINMYAVT	IGGWSRPLVA	
orf53ng-1	FDNNGVYIAS	AVLALVFLAL	GAFVQYNGGE	AVQMAGKGYI	GQLINMYAVT	IGGWSRPLVA	
		160	170	180	190	200	210
		300	310	320	330	340	350
orf53-1.pep	FIAFACMYGT	TIIVVDGYARA	IAEFVRLLR	GDKTKTGNAE	FFAWNIVWAGS	GLAVIFWFDG	
orf53ng-1	FIAFACMYGT	TIIVVDGYARA	IAEFVRLLR	GDKTKTGNAE	FFAWNIVWAGS	GLAVIFWFDG	
		220	230	240	250	260	270
		360	370	380	390	400	410
orf53-1.pep	VMANLLKFAM	IAAFVSAFPV	FAWLNYRLVKG	DEKHKLTSGM	NALALAGL	ITGLTFLVFL	
orf53ng-1	AMAECLKFAM	IAAFVSAFPV	FAWLNYRLVKG	DEKHKLTSGM	NALALAGL	ITGLTFLVFL	
		280	290	300	310	320	330
orf53-1.pep	NLAGMFKX						
orf53ng-1	NLTGLLAX						

Based on this analysis, including the presence of a putative leader sequence (double-underlined) and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be

5 useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 58

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 487>:

```

1  ..TTGGGGGAAA CGGCATATGT TTTGGATAGT TTTGATCGTT ATTTTGTGTG
10  51  TCGCGCTTGCC GGCCTGTTTT TTTGCCGCGC ACAATCCGAA CGCGAGTGGG
    101  TCGCGCGAGT TTTGCGCTGG CAGGAAAGAA AAGGGGAGAA ACAGGCGGAG
    151  CTGCTTGAAA TCAAGACGCG TATGCCGAT TTTCCCGAAC TTGCGCTGAT
    201  GCTTTTCCAC GCGCTCAAAA CGGCACTGTA TTGCTGTTT GTCGCTGTG
    251  TCCGTTTCTG CCGAAACTAT CTGGCGCACG AATCCGAACC GGACAGGCC
    301  GTTCGCGCT..

```

15 This corresponds to the amino acid sequence <SEQ ID 488; ORF58>:

```

1  ..LRETAYVLDS FDRYFVVALA GLFFVRAQSE REWMREVSAA QKKKGKQAE
51  LPEIKDGMPD FPELALMLFH AVKTAVYVLF VGVRFRCRNY LAHESEFDRP
101  VFP..

```

Further work revealed the complete nucleotide sequence <SEQ ID 489>:

```

20  1  ATGTTTTGGA TAGTTTTGAT CGTTATTTTG TTTGCTGCGC TTGCGCGGCTT
    51  GTTTTTTTGTC CGCGCACAAAT CCGAAACGGA GTGGATGGCG GAGGTTTCTG
    101  CGTGGCGAGGA AAAGAAAGGGG GAAAAACAGG CGGAGCTGCC TGAATCAAAA
    151  GACGSTATATG CGATTTTTCG CGAACTTGCC CTGATGCTTT TCCATGCGGT
    201  CAACAACGGCA GTGTATTGCG TGTTTGTGCG TGTGCTGCGT TCTGTCCGAA
    251  ACTATCTGGC GACCGAATCC GAACCGACAC GCGCGTTCG GCGTCTCTC
    301  GCAACCGGTG CGGATGTGCC GAACCGATCC GACGATATT CAGACAGTGG
    351  AAACGGGAGG GAAGAGCGCG AAACCGAAGA AGCAAGCTCT GCGGAGGAAG
    401  AGGCTGCGCA TACGGAAGAC ATTGCAACTG CGGTATCGA CACCGCGCGC
    451  ATCCCATCTG ACCGAGATAT TGCTGAAGGG TTGATGCGCT CTGAAGCGCA
    501  AATTTCGCCC GTCCGTCGCG TTTTAAAGA AATCACTTTG GAAGAAGCAA
    551  CGCGTGCTTT AAACAGCGCG GCTTTAAGGG AAACGAAAAA ACCTATATATC
    601  GATGCATTTG AGAAAAACGA AACAGCGGTC CCCAAGTCC GCGTGTCCGA
    651  TACCCGATG GAGGGCGTGC AGATTATCG TTTGACGAC CCGTGCTTC
    701  AACGCACTGA TTCCCATATG TTGATGCGG ACAAGAAGC GTTTTCGAG
    751  TCTGCGGATT ACGGATTGGA GCGGTATTT GAGAGCGAG ATCCGCTGCG
    801  CTTTTCGCA GTCAAGCGCG AAAATGACAG GAATGCGCG TTCCACGCTC
    851  ATCGAGTCA GGGAAAGGG CAGCGCGAG CAATATCCC GATGTTTCC
    901  CAGGGCAGT CGTTCGAGA CGGACGCGCG CTCGCGATG CCGCGCGCG
    951  CGTTTCGCTC AATTGAAAG AACCGAACAA GCGACGGTT TCTGCGGAGG
    1001  CGCGAATTC TCGCCTGATT CCGGAAAGTC AGACGGTTGT CCGGAAACCG
    1051  GATGTCGAAA TGCGCTCTGA AACCGAAAA TTTTTCACGG AAACCGTTTC
    1101  GTCTGTGGGA TACGCGCGTC CGGTTTATGA TGAACATGCC GATATCCATA
    1151  TTGAAGAACC TGCGCGCGCC GATGCTTGGG TGGTGAACCC ACCGAGGTG
    1201  CGGAAAGTTC CCATGACGCG AATCGATATT CAGCGCGCGC CTCGCGTATC
    1251  GGAAATCTAC AACCGTACCT ATGAACCGCC CTCAGGATT CAGCAGGTGC
    1301  AACGCGCGC CATTCGCGAG ACCGACCATC TTGCGATGA GTTTTGAAT
    1351  GGAGGTTGGC AGGAGGAAC CGCGCTATT GCGGRTGAG CGAGTGAAGG
    1401  TGCGCGGAG CGGTCAAGCG GGCATATCT GTGCGAAACC GAAGCGTTGC
    1451  GGCATGACAG TCAGGCGGTT TGTCCGTTG AAAATGTGCC GTCTGAACGC
    1501  CGCTGCTGCT GCGCTCGGA TACGGAAGCG CAGTACGCG GCTTCCATC
    1551  TGAAAGAACG GTTCGCGTAT CGAACACCTC CGGACACACC GACTGCTTC
    1601  TGCTTCGCT GTTCAATCCC GAGGCGACGC AAACCGAAGA AGACTGTTG
    1651  GAAACAGCA TCACCATCGA AGAAAAATTG GCGGAGTTCA AGTCAAGGT
    1701  CAAGGTTGTC GATTCTTATT CCGCGCCCGT AATTACGGT TATGAATCG
    1751  AACCCGATGT CCGCGTGGCG GGCAATTCC TCTGAATCT GGAAGAAATC
    1801  TTGGCGCGTT CCGCTCGCGT GGCTTCCATC CCGGTTGTC AAACCATCCC
    1851  CGGCAAAACC TGCATGGGTT TGAACCTTC GAACCGAAA CCGCAATGAA

```

1901 TAGCGCTGAG CGAAATCTTC AATTGSCCGG AGTTTGCCGA ATCCAAATCC
 1951 AAGCTGACGC TCGCGCTCGG TCAGGACATC ACCGCGACGC CCGTCGTAAAC
 2001 CGACTTGGGA AAAGCACCGC ATTGTGTGGT TGC CGGCGACG ACCGGTTCGG
 2051 GCAAAATCGGT GGGTGTCAAC GCGATGATTC TGTCTATGCT TTTCAAAGCC
 2101 GCGCCGGAAG ACGTGCCTAT GATTATGATC GATCCGAAAA TGCTGGAATT
 2151 GAGCATTTAC GAAGGCATCC CGCACTCGCT CGCCCTGTCT GTTACCGATA
 2201 TGAAGCTGGC GGCAACGCGG CTGAACCTGG GTTGTAAACG AATGGAAAAA
 2251 CGCTACCGCC TGATGAGCTT TATGGGCGTG CGTAACTCTG CGGGCTTCAA
 2301 TCBAAAATTC GCGCAAGGGG CACCAAGGGG AGAAAAATC GGCATATCGT
 2351 TCAGCTTCAC GCCCGACGAT CCGCAACCTT TGGAAAAACT GCCCTTTATC
 2401 GTGGTGTGGT TCGATGAGTT TGCCGACCTG ATGATGACGG CAGGCAAGAA
 2451 AATCGAAGAA CTGATTGCCG GCCTCGCCCA AAAAGCCGCG GCGCGAGGCA
 2501 TCCATTGATG TCTTGGCCAC CAACGCCCA GCGTCGATGT CATCACGGGT
 2551 CTGATTAAAG CGAACATCCC GACGCGTATC GCGTTCCAAG TGTCACGCAA
 2601 AATCGACAGC CGCACGATTC TCGACCAAAAT GGGCGCGGAA AACCTGCTCG
 2651 GTCAGGCGCA TATGCTGTTC TGCTGCGCGG GTACTGCCTA TCCGCAAGCG
 2701 GTTCACGGCG CGTTTGCTTC GGATGAAGAG GTGCAACGCG TGCTCGAATA
 2751 TTTGAAACAG TTTGGCGAAC CGGACTATGT TGACGATATT TTGAGCGGCG
 2801 GCGGCAAGCA AGAGCTGCCG GGCATCGGCG GCAGCGGCGA CGACGAAACC
 2851 GATCCGATGT ACGACGAGGC CGTATCCGTT GTCTCGTAAA CGCGCAAAGC
 2901 CAGCATTTGC GCGCTACAGC GCGCCTTGGG TATCGGCTAC AACCGCGCGG
 2951 CGCGTCTGAT TGACGAGATG GAGGCGGAG GCATCTGTGT CCGCACCGGA
 3001 CACAACGGCA ACCGTACGAT TCTCTGCCCC TTGGAACATG CTTGA

This corresponds to the amino acid sequence <SEQ ID 490; ORF58-1>:

25 1 MFVIVLIVIL LLALAGLFFV RAQSEREMMR EVSAWQEKKG EKQALPEIK
 51 DGMDFPELA IMLFHAVKTA VYWLFGVVR FCNRYLAHES EPDRPVFPAS
 101 ANRADVPAS DGYSDSGNLT EEAETEEAEA AEEEAADTED IATAIDNRR
 151 IPFDRSIAEG IMPSESEISP VRPFKEITL EEAATRALNSA ALRETKKRYI
 201 DAFEKNETAV PKVRVSDTPM EGLQLIGLDD FVLQRTYSHM FADAKEAFSE
 30 251 SADYGFEPYF EKQHPASAFSA VKAENARNAP FHRHAGOGKG QAEAKSPDVS
 301 QGQSVSDGTA VRDARRRVSV NLKEPNKATV SAEARISRLI PESQTVVGR
 351 DVEMPSETEN VFTETVSVSG YGGFVYDETA DIHIEEPAP DAWVVEPPEV
 401 PKVMTAIDLI QPPFPVSEIY NRTYEPFSGF EOVQSRISAE TDHLADDVLN
 451 GGWQETAAAL ADDGSEGAEE RSGQYLSET EAFHGDSQAV CFENVPSEF
 35 501 PSCHRVSITA DGAFFSEET GAVSEHLFTT DLLLPFLNF EATQTEELL
 551 ENSITIEEKL AEFKVKVKVY DYSVSGVITR YEIEPDVGVVR GNSVIANLEKD
 601 LARSLGVASI RVVETIPGKT CMGLELNPFP ROMIRLSEIF NSPEFAESKS
 651 KLTLALGQDI TGQPVVTDLG KAPHLLVAGT TGSQKSVGVN AMILSMLFKA
 701 APEDVRMIMI DPKMLELSIY EGIPHLLAVR VTDMLAANA LNCVNMEEK
 40 751 RYRIMSMFVG RNLAGFNQKI AEAARGEKI GNFTSLTPD PEPLKLPFI
 801 VVVVDFADL MMTAGKKIEE LIAARACKAR AAGIHILAT QRPVSDVITG
 851 LIKANIPTRI AFOVSSKIDS RTILDQMGAE NLLGQSDMLF LLPGTAYPQR
 901 VHGAFADEE VHRVVEYLKO FGEPTYDDI LSGGSGEELP GIGRSGDDOT
 951 DMYDEAVSV VLKTRKASIS GVORALRIGY NRAARLIDOM EAEGIVSAPE
 45 1001 HNGNRTILVP LDNA*

Computer analysis of this amino acid sequence predicts the indicated transmembrane region, and also gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF58 shows 96.6% identity over a 89aa overlap with an ORF (ORF58a) from strain A of *N.*

meningitidis:

50
 orf58.pep LRETAVLSDSFDYRFVVALAGLFFVRAQSEREMMRVSAWQEKKGKQALPEIKDGMPP
 orf58a MFVIVLIVILLALAGLFFVRAQSEREMMRVSAWQEKKGKQALPEIKDGMPP
 55
 orf58.pep FPELALMLFHAVKTAVYWLFGVVRFCNRYLAHESEPDRPVFP
 orf58a FPELALMLFHAVKTAVYWLFGVVRFCNRYLAHESEPDRPVFPASANRADVPASDGYSD
 60

The complete length ORF58a nucleotide sequence <SEQ ID 491> is:

1	ATGTTTGGGA	TACTTTTGAT	CGTATTTTTG	TTGCTTGGCG	TTGCGCGCTT
51	GTITTTTGTG	CGCGCACAA	CGCAAGCGGA	GTGGATGCGC	GAGGTTTCTG
101	CGTGGCAGGA	AAAGAAGGG	GAAACACAGG	CGGAGCTGCC	TGAARTCAAA
151	GACGGTATGC	CCGATTTTCC	CGACTTTGCC	CTGATGCTTT	TCCATGCCGT
201	CAAAACCGCA	GTGATTGGCG	TGTTTGTGGG	TGTCGTCCGT	TTCTGCCGAA
251	ACTATCTGGC	GCACGAATCC	GAAACGCGCA	GGCCCGTTCC	GCCTGCTTCT
301	GCAAACTCGT	CGGATGTTCC	GACCGCATCC	GACGGATATT	CAGACAGTGG
351	AAACGGGACG	GAAGAAGCGG	AAACGGGAAG	AGCAGAAGCT	CGCGAGGAAG
401	AGGCTGCCGA	TACGGAAGAC	ATTGCAACTG	CCGTAATCGA	CAACCGCCGC
451	ATCCCATTCG	ACCGGAGTAT	TGCTGAAGGG	TTGATGCGCT	CTGAAGCGCA
501	AATTTCGCCG	TCGCTCGGG	TTTTTAAGGA	AATCACTTTG	GAAGAAGCAA
551	CGCGTGCTTT	AAACAGCGCG	GCTTTAAAGG	AAACGAAAAA	ACGCTATATC
601	GATCGATTTC	GAAGAAACGA	AACAGCGCGT	CCCAAGTCCG	CGCTGTCCGA
651	TACCCGATG	GAGGCGCTGC	AGATATTTCG	TTTGGACGAC	CTGTGCTTTC
701	AACGCACTGA	TTCCGTTATG	TTGCAATGCG	ACAAAGAAGC	GTTTTTCGAG
751	TCTGGCGATT	ACGGAATTGA	CGCCTATTTT	GAGAAGCAAC	ATCGTCTCG
801	CTTTTCTGCA	GTCAAAGCGG	AAATGTCCAG	GAATGCGCGC	TTCCGCGCTG
851	ATGCAGGCGA	GGGNAAGGGG	CAGGCGGAGG	CNAATCCCG	GGATGTTTCC
901	CAAGGGCAGT	CCGTTTTCAG	CGGCACAGCC	GTCCGCGATT	CNCGCCGCGG
951	CGTTTCCGCT	AATTTGAAAG	AACGGAACAA	GGCAACGGTT	TCTCGCGAGG
1001	CGCGGATTTT	CGCGCTGATT	CCGGAAAATC	GGACGGTTGT	CGGGAACCGG
1051	GATGTCGAAA	TGCGCTCTGA	AACGGAATAT	GTTTTCACGG	AAANTGTTTC
1101	GTCTGTGGGA	TACGCGGCTC	CGGTTTATGA	TGAAATCGCC	GATATCCATA
1151	TTGAAGAACC	TCCGCGGCCG	GATGCTTGGG	TGGTCCGAAC	ACCCGAAGTG
1201	CCGAAAGTTC	CCATGCCCGC	AATNGATATT	CGCCGCGCGC	CTCCGCTATC
1251	GGAAATCTAC	AACCTACTCT	ATGACACGCG	GGCAGAGTGT	GGCAGAGTGT
1301	AAAGCGGAGG	ACCGATGATC	TTGCGGATGT	TTGCTTTGAT	TTGCTTTGAT
1351	GGAGGTTTGG	AGGAGGAAC	CGCGCTTATT	CGCAATGAAG	CGAATGAGGG
1401	TGTGGCAGAG	CGGTCAAGCG	GGCAATATTT	GTCCGAAACC	GAAGCTTCGG
1451	GGCATGACAG	TCAGGCGGTT	TGTCCGTTTG	AAAATGTGCC	GTCTGAACGC
1501	CCGTCGCGCC	GGGCATNGGA	TACGGAAGCG	GATGAAGGGG	GGTTTCCAAT
1551	TGAAGAACC	GTTGCGGTAT	CCGAACACCT	CGCCACAACG	GACCTGCTTC
1601	TGCCCGCGCT	GTTCAATCCC	GGGGCGAGCG	AAACCGAAGA	AGANCTGTTG
1651	GANAACAGCA	TCACCATCGA	AGAAAAATNG	CGCGAGTTCA	AAGTCAAGGT
1701	CAAGGTTTGC	GATTCTTATT	CCGCGCCGCT	GATTACGCGT	TATGAATTCG
1751	AACCCGATGT	CGGCGTGCGC	GGCAATTCGG	TTCTAAATCT	GGAAAAAGAN
1801	TTGGCGCGTT	CGCTCGGCGT	GGCTTCCATC	CGCGTTGTGC	AAACCATCTC
1851	CGGCCAAACC	TGTTATGGTT	TGGAATCTTC	GACCCCGAAA	CGCCAAATGA
1901	TACGCTCAG	CGAATCTTCT	AATTCCCGCC	AGTTTCCGCA	ATCCAAATTC
1951	AAGCTGACGC	TCCGCGTCCG	TCAGGCATCT	ACCGAGACGC	ACCGGTATAC
2001	CGACTTGGGC	AAAGCACCCG	ATTTTGTGGT	TGCCGCGACG	ACCGTTTCGG
2051	GCAAACTCGT	GGGTGTCAAC	CGCATGATTC	TGCTATGCT	TTTCAAAGCC
2101	CGCCCGGAAG	ACGTTGCGTAT	GATTATGATC	GATCCGAAAA	TGCTGGAATT
2151	GAGCATTTAC	GAAAGCATCC	CGCACTTGCT	CGCCCTTGAT	GTTACCGATA
2201	TGAAGCTGGC	GGCAACCGCG	CTGAATCTGT	GTGTTAAACG	AATGGAAGAA
2251	CGCTACCGCC	TGATGAGCTT	TATGGCGCTG	CGCAATCTTG	CGGNTTCAA
2301	TCAAAAATC	CGCGAGCGCG	CAGCAAGGGG	GGAGAAAAAT	GGCAACCCGT
2351	TCAGCGCTC	CGCGCAACAT	CCCGAACCTT	TGGANAAATT	CGCGTTTATC
2401	GTGTCGTGG	TTGATGAGTT	TGCCGACCTG	ATGATGACGG	CAGGCAAGAA
2451	AATCGAAGAA	CTGATTGCCG	CGCTCGCCCA	AAAAGCCCGC	GGGCGAGGCA
2501	TCCATCTTAT	CTTGCCACCA	CACACGCCCA	GTGTCGATGT	GATCAACGGT
2551	CTGATTAGG	CGACATCTCC	GACGCTATTC	GGCTTCCAG	GTGTCGACGA
2601	AATCGACAGC	CGACAGCATTC	TTGACCAATAT	GGGTGCGGAT	AACTGCTCTG
2651	GGCAGGGCGA	TATGCTGTTT	CTGCCGCGGG	GTACGSCCTA	TCCGACGCGC
2701	GTTCAAGCGG	CGTTTGCCCT	GGAATGAAGAG	GTGCACGCGC	TGTCGGAATA
2751	TCTGAAACAG	TTTGGCGAAG	CGGCAATATG	TGACGATATN	TTGAGCGGGC
2801	GTATGTCCGA	CGATTTGCTG	GGAATCAGCC	GGAGCGGGCA	CGGCGAAGCC
2851	GATCCGATGT	ACGACGAGGC	CGTGTGNTTT	GTTTTGAAAA	CGCGCAAGGC
2901	CAGCATTTCT	GGCGTGCAGC	GCGCAATTGC	TATCGCTGCT	AATCGCGCGC
2951	CGCGCTGAT	TGACCAGATG	GAGCGGAAG	CAGTTGTGTC	CGACCGGGAA
3001	CACAAACGGA	ACCGTACGAT	TCTGTCGCC	TNGACAATG	CTTGA

This encodes a protein having amino acid sequence <SEQ ID 492>:

1	MFWIVLVIL	LLALAGLFFV	RAQSEREWNR	EVSAAWEKKG	EKQALPEIK
51	QGMFDPEELA	LMLEHAKVTA	VYLFVGVVVR	FCRNYLAREH	EPDREVPFAS
101	ANRADYFPEH	QVYSQSGNTG	EEAFETAEAA	AEBAADTEI	INAVINDNR
151	IFPDRLAEG	LMPSSEISLP	VRFVEKEITL	EEATRLANSA	ALAEIKKRYI
201	DAFEKNETAV	PKVRVSDTFM	EGLQIIGLDD	PVLQRTYSRM	FDADKEAFSE

251	SADYGFEPYF	EKQHPSAFSA	VKAENARNAP	FRRHAGQGKQ	QAEAKSPDVS
301	QGQSVSDGTA	VRDAXRRVSV	NLKEPNKATV	SAEARI SRLI	PESRTVVGKR
351	DVEMPSETEN	VTEKXVSVSG	YGXPFVYDETA	DIHIEEPAAP	WDAMVVSPPVEV
401	PKVPMAPXDI	PPPPPVSEIY	NRTYEPAPAG	EQVQSRIRAE	TDHLADDVLN
451	GGWQZETAAL	ANDGSEGVAE	RSSQYLSET	EAFGHDSQAV	CPFENVPSE
501	PSRRAXDTEA	DEGAFQSEET	GAVSEHLPTT	DLLLPPLFNP	GATQTEKXLL
551	XNSITIEEKX	AEKFKRVKVV	DSYSGPVITR	YEIEPDVGVK	GNSVLNLEKX
601	LARSLGVASI	RVETILGKT	CMLELFPNK	RQMIRLSEIF	NSPEFASKS
651	KILTLALGQD	TGQPVVTDG	KAPHLILACT	TGSKSGVGVN	AMILNMLFKA
701	APEDVRMIMI	DPKMLSLISY	EGIPHLLAPV	VTONKLAANA	LNCVNMEEK
751	RYRLMSFMVG	RNLAGNKNQI	ASAAAGSEKI	GNPFSLTPEP	PEPLXKLPEI
801	VVVVDEFAFL	MMTAGKKIEE	LIARLAQAKR	AAGIHILILAT	ORPSVDVITG
851	LKANIPTRI	AFOVSSKIDS	RTILDQMGAE	NLLGQDMLF	LPPGTAYPQR
901	VHGAFASDEE	VHRVVSYLKQ	FGEPDYVDX	LSGGMSSDLL	GISRSQDGET
951	DPMYDEAVSV	VLKTRKASIS	GVQRALRIGY	NRAARLDQM	EAGGIVSAPE
1001	HNGNRTILVP	XDNA*			

ORF58a and ORF58-1 show 96.6% identity in 1014 aa overlap:

		10	20	30	40	50	60
20	orf58a.pep	MFWIVLIVILLALAGLFFVRAQSEREMREVS	SAWQSKGKQAE	PEIKDGMDFPELA			
	orf58-1	MFWIVLIVILLALAGLFFVRAQSEREMREVS	SAWQSKGKQAE	PEIKDGMDFPELA			
		10	20	30	40	50	60
25	orf58a.pep	70	80	90	100	110	120
	orf58-1	LMLFHAVKTAVYWL	FVGVVRF	CRNYLAHSE	PPORVPVPPASANRADVPTASDGYSDSGNGT		
		70	80	90	100	110	120
30	orf58a.pep	130	140	150	160	170	180
	orf58-1	EEAETEEAEAEAEAAADTE	DIATAVIDNRRIP	FDORSIASGLMPSSSEIS	SPVRPVFKEITL		
		130	140	150	160	170	180
35	orf58a.pep	190	200	210	220	230	240
	orf58-1	EEATRALNSAALRET	KKRYIDAF	EKNETA	VPKRVSDT	PMSGLQII	GLDPPVLQRTYSRM
		190	200	210	220	230	240
40	orf58a.pep	250	260	270	280	290	300
	orf58-1	FDADKEAFSESADYGF	EPYFEKQHP	SFAVKAENARNAP	FRRHAGQGKQ	QAEAKSPDVS	
		250	260	270	280	290	300
45	orf58a.pep	310	320	330	340	350	360
	orf58-1	QGQSVSDGTA	VRDAXRRVSV	NLKEPNKATV	SAEARI SRLI	PESRTVVGKR	DVEMPSETEN
		310	320	330	340	350	360
50	orf58a.pep	370	380	390	400	410	420
	orf58-1	VTEKXVSVGYG	KXPFVYD	ETADIHIEE	PAAPDAWV	PEPEVKVPM	PAXDI PPPPVSEIY
		370	380	390	400	410	420
55	orf58a.pep	430	440	450	460	470	480
	orf58-1	NRTYEPAPAG	EQVQSRIRAE	TDHLLADDVINGWQ	ZEETAALANDGSEGA	VERSSQYLS	ET
		430	440	450	460	470	480
60	orf58a.pep	490	500	510	520	530	540
	orf58-1	EAFGHDSQAV	CPFENVPSE	RPSRRAXDTE	ADGAFQSEET	GAVSEHLPTT	DLLLPPLFNP
		490	500	510	520	530	540
70	orf58a.pep	490	500	510	520	530	540
	orf58-1	EAFGHDSQAV	CPFENVPSE	RPSRRAXDTE	ADGAFQSEET	GAVSEHLPTT	DLLLPPLFNP

-290-

		550	560	570	580	590	600
	orf58a.pep	GATQTEEXLLXNSITIEEKXAEFKVKVVDYSYSGPVITRYEIEPDVGVGRGNSVNLKEX					
5	orf58-1	EATQTEELLNSITIEEKLAEFKVKVVDYSYSGPVITRYEIEPDVGVGRGNSVNLKED					
		550	560	570	580	590	600
	orf58a.pep	LARSLGVASIRUVETILGKTCMGLLPNPKQMIRLSEIFNSPEFAESKSKLTALGQDI					
10	orf58-1	LARSLGVASIRUVETIPGKTCMGLLPNPKQMIRLSEIFNSPEFAESKSKLTALGQDI					
		610	620	630	640	650	660
	orf58a.pep	TGQPVVTDLGKAPHLLVAGTTGSGKSGVGNAMILSMLEKAAPEVDVRMIMIDPKMLELSIY					
15	orf58-1	TGQPVVTDLGKAPHLLVAGTTGSGKSGVGNAMILSMLEKAAPEVDVRMIMIDPKMLELSIY					
		670	680	690	700	710	720
	orf58a.pep	EGIPHLLAPVVTDMKLAANALNWCNEMEKRYRLMSFMGVNLAGXNOKIAEAAARGEKI					
20	orf58-1	EGIPHLLAPVVTDMKLAANALNWCNEMEKRYRLMSFMGVNLAGXNOKIAEAAARGEKI					
		730	740	750	760	770	780
	orf58a.pep	GNPFSLTPDNFEPLKLPFIVVVVDEADLMMTAGKKIEELIARLAQKARAAGIHLILAT					
25	orf58-1	GNPFSLTPDDEFEPLKLPFIVVVVDEADLMMTAGKKIEELIARLAQKARAAGIHLILAT					
		790	800	810	820	830	840
	orf58a.pep	ORPSVDVITGLIKANIPTRIAFPQVSSKIDSRITLQDMGAENLLGGGDMLEFLPGTAYFQR					
30	orf58-1	ORPSVDVITGLIKANIPTRIAFPQVSSKIDSRITLQDMGAENLLGGGDMLEFLPGTAYFQR					
		850	860	870	880	890	900
	orf58a.pep	VHGAFAASDEEVRVVEYLKQFGEPPYVDLISGGGSEELPGIGRSGGDETDPMYDEAVSV					
40	orf58-1	VHGAFAASDEEVRVVEYLKQFGEPPYVDLISGGGSEELPGIGRSGGDETDPMYDEAVSV					
		910	920	930	940	950	960
	orf58a.pep	VLKTRKASISGVQRALAIQYNRAARLIDQMEAEAGIVSAPEHNGNRTILVFXDNAX					
45	orf58-1	VLKTRKASISGVQRALAIQYNRAARLIDQMEAEAGIVSAPEHNGNRTILVFXDNAX					
		970	980	990	1000	1010	

50 Homology with a predicted ORF from *N.gonorrhoeae*

ORF58 shows complete identity over a 9aa overlap with a predicted ORF (ORF58ng) from *N.gonorrhoeae*:

	orf58.pep	ALMLPHAVKTAAYVWLVGVVVRFCRNYLAHSEPDAPVFP	103
55	orf58ng	SEPDAPVFPASANRADVPTASDGYSDSNG	30

The ORF58ng nucleotide sequence <SEQ ID 493> is predicted to encode a protein having partial amino acid sequence <SEQ ID 494>:

	1	..SEPDAPVFP	SANRADVPTA	SDGYSDSNG	TEEAETAAE	AAEEEAADTE
	51	DIATAVIDNR	RIPDFRSIAE	GLMQSEKTS	PVRPVFKEIT	LEEATRALSS
60	101	AALRETKKRY	IDAFENKNGTA	VFKVRSVDT	MEGLQIIGLD	DPVLQRTYSR
	151	MFDAKEAFS	ESADYGFEPY	FEKQHPASFS	AVKAENARNA	PFRHAGQEK
	201	GQAEAKSPDV	SOQGSVSDGT	AVRDARRRS	VNLKEPNKAT	VSAAERISRL
	251	IPESRTVVGK	RDVENFSETE	NVFTETVSSV	GYGGPVYDEA	ADHIEEPAA
	301	PDAWVVEPPE	VPEVAVFEID	ILPPPPVSEI	YNRTYEPGAP	FEQAQRSRJA

351	ETDHLAADVL NGGWCEETA	IADDGSEGAA	ERSSGQYLSE	TEAFGHDSQA
401	VCFPFEDVSE RFSRCVSDTE	ADEGAFQSEE	TGAVSEHLPT	TDLLPLPLFN
451	PEATQTEEL LENSITIEEK	LAEFVKVKV	VDSYSGPVIT	RYEIEPOVGV
501	RGNISVLNLEK DLARSILGVA	IRVVETIPGK	TCMGLELNF	KRCMIRLSEI
551	FNSPEFAESK SKLTALGQD	ITGQPVVTDL	GKAPHLVAG	TTGSGKSVGV
601	NAMILSMLEK AAFEDVRMIM	IDPKMLELSI	YEGITHLLA	VVTDMKLAAN
651	ALNWCVNEME KRYRLMSFMG	VRNLAGEFNK	IAEAAARGEK	IGNPFSLTDP
701	DDEPELEKLEF IVVVVDEPAD	LMMTAGKKIE	ELIARLAQKA	RAAGHLILA
751	TCRFSVDVIT GLIKANIPTR	IAFQVSSKID	SRTLLDQMG	ENILGQSGDML
801	FLPFGVAVFP RVHGAQAGE	FVHRVVEYLL	QFGEFDVDD	ILSGGSGSEL
851	PGIGRSGDGE TDRMYDEAVS	VVLKTRKASI	SGVQRLALIG	YNRAARLIDQ
901	MEAEIGVISAP EHNHRTILV	PLDNA		

- This partial gonococcal sequence contains a predicted transmembrane region and a predicted ATP/GTP-binding site motif A (P-loop; double underlined). Furthermore, it has a domain homologous to the FTSK cell division protein of *E. coli*. Alignment of ORF58ng and FtsK (accession number p46889) show a 65 % amino acid identity in 459 overlap:

ORF58ng:	467	IEEKLAEFKVKVKVDSYSGPVITRYEIEPDVGRGNSVLNLEKDLARSILGVAIRVVET	526
FtsK:	868	VEARLADPRKADVNVYSPGPVITRPFELNLAGVKAARISNLSDLASLSTVAVRVVEV	927
ORF58ng:	527	IPGKTCMGLELFPNPKRQMRLEISFNSPEFAESKSKLTALGQDITGQPVVTDLGAHPL	586
FtsK:	928	IPGKFPVYGLFPNKKRQTVYLRVLDNAKFRDNPSFLTIVVLGKDIAQEPVADLAKMHPL	987
ORF58ng:	587	LVAGTTGSGKSVGNAMILSMLFKAAPEDVRMIMIDPKMLELSIYEGITHLLAPVVTOMK	646
FtsK:	988	LVAGTTGSGKSVGNAMILSMLFKAAPEDVRMIMIDPKMLELSIYEGITHLLAPVVTOMK	1047
ORF58ng:	647	LAANALNWCVNEMEKRYRLMSFMGVRNLAGEFNKIAEAAARGEKIGNPFSLTDPDEP--	704
FtsK:	1048	DAANALNWCVNEMERRVKMSALGVRNLAGEFNKIAEADRMRRPFDIPYKNGKSGDMDAQH	1107
ORF58ng:	705	--LEKLPFIVVVVDEPADLMMTAGKKIEELIARLAQKARAAGHILATCRPSVDVITGL	762
FtsK:	1108	FLVKKEPYIVLVDEPADLMMTAGKKIEELIARLAQKARAAGHILATCRPSVDVITGL	1167
ORF58ng:	763	IKANIPTRIAFQVSSKIDSRITLDQMGAEENLLGQGDMLFLPFGTAYPQRVHGAFAFASDEEV	822
FtsK:	1168	IKANIPTRIAFVSSKIDSRITLDQGAELLLGQGDMLFLPFGTAYPQRVHGAFAFVDRQEV	1227
ORF58ng:	823	HRVVEYLYKGFGEFDVDDILSGGSEELPGIGRSGDGETDPMYDEAVSVLTKRASKISG	882
FtsK:	1228	HVVQDWKARGRPQVVDGITSDESEGGAG-GFDGAELDPLFDQAVQFTEKKRASKISG	1286
ORF58ng:	883	VQRALRIGYNRAARLIDQMEAEIGVISAP EHNHRTILV	921
FtsK:	1287	VQRFRIGYNRAARLIDQMEAEIGVISAP EHNHRTILV	1325

Further work on ORF58ng revealed the complete gonococcal DNA sequence to be <SEQ ID 495>:

1	ATGTTTGGGA TAGTTTGTAT CGTTATgtg TTGCTTGCGC TTGCCGCGCT
51	GTTTGTGTC CGCGCACAA CCGAACCGGA GTGGATGCGC GAGGTTCTTG
101	CGTGGCAGGA AAGAAAGAGG GAAAAACAGG CGGAGCTGCC TGAATACAAA
151	GACGGTATGC CCGATTTC CCGATTTC CCGATTTC TCCATCGCGT
201	CAAAAACGGCA GTGATTGGC TGTTTGTGCG TGCTCGCTCG TTCTGCGGAA
251	ACTATCTGCG GCACGAATCC GAACCGGACA GGCCCGTTCC GCTCTGCTTC
301	CAACACCGTG CGSATTGTCC GACCGCATCC GACGGGTATT GACAGCATGG
351	AAACGGGACG GAAAGACGCG AACCGAGCTG ACAGAGACTG CGGAGAGAGG
401	AGCTCCCGCA TACGGAAGAC ATTGCATCTG CCGTAACTCG CAACCGCGCG
451	ATCCcatTCG ACCGGAGTAT TGCTGAAGGG TTGATGCATG CTGAAGACCA
501	AACCTGCGCC GTCCGTCCGG TTTTAAAGGA AATCACTTTG GAAGAAGCAA
551	CGCGTGCTTT AAGCAGCGCG GCTTTAAGGG AAACGAAATA ACCTATATCG
601	GATGCATTAT AGAAAAACGG AACAGCGCTC CCCAAGATAC CGGTGCTCGA
651	TACCCCGATG GAAGGGCTGC AGATTATCGG TTGAGACGAC CCGTGCTTC
701	AACGCAAGTA TTCCGATATG TTGATGCGGG ACAAGAGAAC GTTTCCGAG
751	CTGCGGATT ACAGATTGGA GCGGATTTT GAGAAGCAGC ATCCGCTCG

5	801	CTTTTCTGCA	GTCAAAGCCG	AAAATGCAAG	GAATGCGCCG	TTCCGCCGTC
	851	ATCGAGGGCA	GGAGAAAGGG	CAGCGGAGAG	CAAAATCCCC	GGATGTTTCC
	901	CAAGGGCAGT	CGGTTTTCAGA	CGGCACAGCC	GTCGCGGATG	CCCGCGCCGC
	951	CGTTTCCGTC	AATTTGAAAG	AACCGAACAA	GGCAACGGTT	TCTCGGGAGG
	1001	CGCGGATTT	CGCGCTGATT	CCGGAAGATC	GGACGGTTGT	CGGGAAACCG
	1051	GATGTCGAAA	TGCCGTCTGA	AACCGAAAAT	GTTTTCACGG	AAACCGGTTT
	1101	GTCCTGTGGG	TACGGCGGTC	CGGTTTATGA	TGAAGCTGCC	GATATCCATA
	1151	TTGAAGAGCC	TGCGCGCCCC	GATGCTTGGG	TGCTCGAAC	ACCCGAAGTG
	1201	CGGAGGTG	CGCTACGCA	AATCGATGCA	GTCCGCGCCG	CTCCCGTATG
10	1251	GAAATCTAC	AACCGTTATC	ATGAGCGCCG	GGCAGGATTC	GAGCAGGCTC
	1301	AACGCAGCCG	CATTGCGGAA	ACCGACATC	TTGCGCGTGA	TGTTTGTAAAT
	1351	GGAGGTTGCG	AGGAGGAAAC	CGCCGCTATT	CGAGATGACG	CGAGTGAGGG
	1401	TGCGGCAGAG	CGGTCAAGCG	GGCAATATCT	GTCGGAACCC	GAAGCGTTCG
	1451	GGCATGACAG	TACGGCGGTT	TGTCGCTTGG	AAGATGTGCC	GTCTGAACGC
15	1501	CGGTCTGCG	GGGTATCGGA	TACGGAAGCG	GATGAAGGGG	CGTTCGAATC
	1551	GGAAAGAGCC	GGTGGGGTAT	CGGAACACCT	GCCGACAAAC	GACCTGCTTC
	1601	TGCCCTCGCT	GTTCAATCCC	GAGGCGACGC	AAACCGAAGA	AGAACTGTTG
	1651	GAAGAACGCA	TCACCATCGA	AGAAAAATTG	CGCGGATCCA	AAGTCAAGGT
20	1701	CAAGGTTGTC	GATTCTTATT	CGCGCCCGCT	GATTACGGGT	TATGAAATCG
	1751	AACCCGATGT	CGCGCTGCGC	GGCAATTCG	TTCTGAATT	GGAAAAAGAC
	1801	TTGGCGGCTT	CGCTCGGCGT	GGCTTCCATC	CGGCTTTGCG	AAACCATCCC
	1851	CGCGAATACC	TGCTATGGCT	TGGAACCTTC	GAACCCGAAA	CGCCGATATG
	1901	TACGCGTGA	CGAATTTTCC	ATTGCGCGCG	AGTTTTCGGA	ATAAATCCC
	1951	AAGCTGACCG	TGCGCTCTGG	TCAGACATTC	ACCGGACAGC	ACGCTCTAAC
25	2001	CGACTTGGGC	AAAGCACCGC	ATTTCGTGGT	TGCGCGGACG	ACCGGTTCCG
	2051	GCAATCGGT	GGGTGTCAAC	GCGATGATTC	TGCTATGCTC	TTTCAAAGCC
	2101	GCGCGGGAAG	ACGTGCGTAT	GATTATGATC	GATCCGAAAA	TGCTGAAGTT
	2151	GAGCATTTAC	GAAGGCATCA	CGCACTGCTC	CGCCCTGTGC	GTTACCGATA
	2201	TGAAGCTGGC	GGCAAAACGC	CTGAACCTGT	GTGTTAACGA	AATGGAAAAA
30	2251	CGCTACCGCC	TGATGAGCTT	TATGGGCGTG	CGCAATCTTG	CGGGCTTCAA
	2301	CCAAAAAATC	GCGGAAGCCG	CAGCAAGGGG	AGAAAAATTC	GGCAATCCGT
	2351	TCAGGCTCAC	GCGCCAGCAT	CCGGAACCTT	TGGAAAAATC	GCGGTTTATC
	2401	GTGGTCTGGG	TGCGATGAGTT	TGCGGATTTG	ATGATGACGG	CAGGCAAGAA
35	2451	AATCGAAGAA	CTGATTCGCG	GCTCTGCCCA	AAAGCCCGCG	CGGCGAGGCA
	2501	TCGACCTTAT	CTGTCACACA	CAGCCGCCCA	CGCTCGATGT	CATCACGGGT
	2551	CTGATTTAAG	CGAACATCCC	GAGCGCTATC	CGGTTCCGAG	TGTCGAGCAA
	2601	AATCGACAGC	CGCACGATTC	TCGACCAAAT	GGGCGCGGAA	AACTCTGCTC
	2651	GTCAGGGGGA	TATGCTGTTC	CTGCGCGCGG	GTACTGCCTA	TCGCGACGCG
40	2701	GTTACAGGCG	CGTTTGCTCT	GGATGAAGAG	GTGACACGGG	TGGTCCAATA
	2751	TCGGAAGCAG	TTTGCGGAGC	CGGACTATGT	TGACGATATT	TTGAGCGGCG
	2801	GCGGCAGCGA	AGAGCTGCC	GGCATCGGGC	GCAAGCGGGA	CGGCGAAACC
	2851	GATCCGATGT	ACGACGAGGC	CGTATCGGTT	GTCTGTAATA	CGGCGAAAGC
	2901	CAGCATTTTG	GGGTACAGCG	GCGCCTTGCG	CATCGGCTAC	AACCGCGCCG
	2951	CGGCTCTGAT	TGACCAAAAT	GAAAGCGAAG	CGATTGTGTC	CGCACCGGAA
45	3001	CACAACGGCA	ACGTCAGCAT	TCTCGTCCCT	TTGGACAATG	CTTGA

This corresponds to the amino acid sequence <SEQ ID 496; ORF58ng-1>:

50	1	MFIVILVIV	LLALAGLFEV	RAQSEREMMR	EVSAWQKKKG	EKQALFEIK
	51	DGMDFPEFS	LMFHAVKTA	VYMLFVGVVR	PCNVLAHES	FDFRVFPAS
	101	ANRADVPAS	DGYSDSGNGT	EEAFTEAAEA	AEFEAAATED	IATVIDNRNR
	151	IPFDRSIAEG	LMQSESKTSP	VRVPFKETLT	EEATRALISA	ALRETKKRYI
	201	DAFEKNGTAV	PKVVRSDTME	EGLQIIGLDD	PVLQRTYSRM	FADKAEAFSE
	251	SADYGFPEFY	EKQHPSAFSA	VKAENARNAP	FRRHAGQKQG	QAEAKSPDVS
	301	QGQSVSDGTA	VRDARRRVSV	NLKEPNKATV	SAEARISRLI	PESTRVUVGKR
	351	DVEMPSETEN	VFTETVSSVG	YGGPVYDEAA	DIHIEEPAP	DAWVVEPPEV
55	401	FEVAVPEIDI	LPFPVSEIY	NRTYEFPAG	EQAQRSRIAE	TDHLAADVLN
	451	GGWQEEATAI	ADDSGEGAAE	RSSGQYLSET	EAFGDSQAV	CFPFDVPSER
	501	PSCRVSDTEA	DEGAFQSEET	GAUHEHLPTT	DLLLPLPLFN	EATQTEEELL
	551	ENSITIEEKL	AEFKVKVKVY	DSYSGFVITR	YEIEPDVGVR	GNSVNLLEKD
	601	LARSLGVASI	RVVETIPGKT	CMGLELNPKN	RQMIRLSEIF	NSPEFAESKS
60	651	KLTLALGGDI	TGQPVVTDLG	KAPHLVAGAT	TSGSGKSVGN	AMILSLMFLA
	701	AFEDVMIMH	DEMLLELSIV	EGITHLAFV	VTVMKLAAAN	LWNCVNEMEK
	751	EYRLMSFMV	RNLACTNCKI	AEAAARGEKI	GNFSLFPDD	PEPELEKPEI
	801	VYVVDPEADL	MMTAGCKTIEF	LIALLAQKAR	AAQIHLILAT	QRFSDVITFG
	851	LTKANIPTRI	AFQVSSKIDS	RTILDQMGAE	NLLGQSDMLF	LPFGTAYFOR
65	901	VHGAFAFDEE	VHRVVEYLKQ	FGFEDYVDLI	LSGGGSEELP	GIGRSGDGET
	951	DEMYDEAVSV	VLKTRKASIS	GVQRLRIGV	NRAARLDQM	EAEIGVSAFE
	1001	HNGNRTILVE	LDNA*			

ORF58ng-1 and ORF58-1 show 97.2% identity in 1014 aa overlap:

		10	20	30	40	50	60
	orf58-1.pep	MFWIVLIVILLALLALAGLFFVRAQSEREMREVSAAQKKGEKQAELEPTKDGMPDFPELA					
5	orf58ng-1	MFWIVLIVILLALLALAGLFFVRAQSEREMREVSAAQKKGEKQAELEPTKDGMPDFPEFS					
		10	20	30	40	50	60
	orf58-1.pep	70	80	90	100	110	120
	orf58ng-1	LMLFHAVKTAVYWLFGVVRFCRNYLAHESEPDRPVPPASANRADVPTASDGYSDSNGGT					
10		70	80	90	100	110	120
	orf58-1.pep	130	140	150	160	170	180
15	orf58ng-1	EEAETEEAAAEAAADTEDIATAVIDNRRIPFDRSIAEGLMPSESEISFVRPVFKEITL					
		130	140	150	160	170	180
	orf58-1.pep	190	200	210	220	230	240
20	orf58ng-1	EEATRALSAALARETKKRYIDAFERKNGTAVPKVVSVDTEMEGLQIIGLDDPVLQRTYSRM					
		190	200	210	220	230	240
25	orf58-1.pep	250	260	270	280	290	300
	orf58ng-1	FDADKEAFSESADYGFEPYFEKQHPSAFSAVKAENARNAPFRRHAGQGGKQAEKASPDVS					
30		250	260	270	280	290	300
	orf58-1.pep	310	320	330	340	350	360
35	orf58ng-1	QGQSVSDGTAVRDARRRVSNNLKEPNKATVSAEARSILILIPESQTVVGKRDVEMPSETEN					
		310	320	330	340	350	360
	orf58-1.pep	370	380	390	400	410	420
40	orf58ng-1	VFTETVSSVGYGGPVYDEADIIIEEPAAPDAWVVEPVEPVKPMTAIDIQPPFPVSEIY					
		370	380	390	400	410	420
45	orf58-1.pep	430	440	450	460	470	480
	orf58ng-1	NRTYEPPSGFEQVQRSAIAETDHLAADVINGGWQSEETAIAADGGSEGAERSSQGYLSET					
		430	440	450	460	470	480
50	orf58-1.pep	490	500	510	520	530	540
	orf58ng-1	EAFGHDSQAVCPFENVPSESRPSRVSDTEADEGAFPSSEETGAVSEHLPTTDLPLPLFNP					
		490	500	510	520	530	540
55	orf58-1.pep	550	560	570	580	590	600
	orf58ng-1	EATQTEELLENSITIEKLAEPKVKVVDVSDSYSGPVITRYEIEPDVGVGRNSVNLKED					
60		550	560	570	580	590	600
	orf58-1.pep	610	620	630	640	650	660
	orf58ng-1	LARSLGVASIRVVETIPGKTCMGLELNPVKQIRLSEIFNSPEFAESKSKILTALQODI					
65		610	620	630	640	650	660
	orf58-1.pep	670	680	690	700	710	720
70	orf58ng-1	TGQPVVTDLGKAFHLLVAGTTGSGKSVGVNAMILSMLFKAAPEDVRKIMIDPKMELSIY					
		670	680	690	700	710	720

		730	740	750	760	770	780
	orf58-1.pep	EGIPHLLAPVVTDMKLAANLNCVNMCKRYRLMSFMGVRNLAFGNQKIAEAAARGEKI					
5	orf58ng-1	EGITHLLAPVVTDMKLAANLNCVNMCKRYRLMSFMGVRNLAFGNQKIAEAAARGEKI					
		730	740	750	760	770	780
	orf58-1.pep	GNPFSLTPODPEPLEKLPPIVVVVDEPADLMMTAGKKIEELIARLAQKARAAGIHLILAT					
10	orf58ng-1	GNPFSLTPODPEPLEKLPPIVVVVDEPADLMMTAGKKIEELIARLAQKARAAGIHLILAT					
		790	800	810	820	830	840
	orf58-1.pep	GNPFSLTPODPEPLEKLPPIVVVVDEPADLMMTAGKKIEELIARLAQKARAAGIHLILAT					
	orf58ng-1	GNPFSLTPODPEPLEKLPPIVVVVDEPADLMMTAGKKIEELIARLAQKARAAGIHLILAT					
		790	800	810	820	830	840
	orf58-1.pep	ORPSVDVITGLIKANIPTRIAFQVSSKIDSRITLQDMGAENLLGGQDMFLPLPGTAYPQR					
15	orf58ng-1	ORPSVDVITGLIKANIPTRIAFQVSSKIDSRITLQDMGAENLLGGQDMFLPLPGTAYPQR					
		850	860	870	880	890	900
	orf58-1.pep	ORPSVDVITGLIKANIPTRIAFQVSSKIDSRITLQDMGAENLLGGQDMFLPLPGTAYPQR					
	orf58ng-1	ORPSVDVITGLIKANIPTRIAFQVSSKIDSRITLQDMGAENLLGGQDMFLPLPGTAYPQR					
		850	860	870	880	890	900
	orf58-1.pep	VHGAFASDEEVHRVVEYLKQFGPEPDYVDDILSGGGSEELPGIGRSGDGETDPMYDEAVSV					
20	orf58ng-1	VHGAFASDEEVHRVVEYLKQFGPEPDYVDDILSGGGSEELPGIGRSGDGETDPMYDEAVSV					
		910	920	930	940	950	960
	orf58-1.pep	VHGAFASDEEVHRVVEYLKQFGPEPDYVDDILSGGGSEELPGIGRSGDGETDPMYDEAVSV					
	orf58ng-1	VHGAFASDEEVHRVVEYLKQFGPEPDYVDDILSGGGSEELPGIGRSGDGETDPMYDEAVSV					
		910	920	930	940	950	960
	orf58-1.pep	VLKTRKASISGVQORALRIGYNRAARLIDQMEAGIVSAPEHNGNRTILVPLDNAX					
25	orf58ng-1	VLKTRKASISGVQORALRIGYNRAARLIDQMEAGIVSAPEHNGNRTILVPLDNAX					
		970	980	990	1000	1010	
	orf58-1.pep	VLKTRKASISGVQORALRIGYNRAARLIDQMEAGIVSAPEHNGNRTILVPLDNAX					
	orf58ng-1	VLKTRKASISGVQORALRIGYNRAARLIDQMEAGIVSAPEHNGNRTILVPLDNAX					
		970	980	990	1000	1010	
30	Furthermore, ORF58ng-1 shows significant homology to the <i>E.coli</i> protein FtsK:						
	sp P46889 FtsK_ECOLI CELL DIVISION PROTEIN FtsK >gi 1651412 gnl PID d1015290 (D1 division protein FtsK [Escherichia coli]) >gi 1651418 gnl PID d1015296 (D90727) Cell division protein FtsK [Escherichia coli] Length = 1329						
35	Score = 576 bits (1469), Expect = e-163						
	Identities = 301/459 (65%), Positives = 353/459 (76%), Gaps = 5/459 (1%)						
	Query: 556	IEEKLAEFKVKVVDYSYSGPVITRYEIEPDVGVGNVNLNLEKDLARSLGVASIRVET	615				
		+E +LA+F++K VV+ GPVITR+2+ GV+ +NL +DLARSL ++RVVE					
40	Subject: 868	VEARLADPRIKADVNNYSPGPVITRFEINLAPGVKAAIRINLSRDLARSLSTVAVRVVE	927				
	Query: 616	IPGK+CMGLELNPVKRQMRILSEIFNSPEFAESKSLTLALGQDITGGPVVTLGKAPHL	675				
		IPGK +GLELNP KRQ +L E+ ++ +F +S LT+ LG+DI +GPVV DL K PHL					
	Subject: 928	IPGKPYVGLLELNPVKRQVYLVLEVDNAKFRNPSPLTVVLGQDIAGEFVVDALAKMPLH	987				
45	Query: 676	LVAGTTGSGKSVGVNAMILSMLFKAAPEDVRMIMIDPKMLELSIYSGITHLLAPVVTDMK	735				
		LVAGTTGSGKSVGVNAMILSML+KA PEDVR IMIDPKMLELS+YSGI HLL VVTDMK					
	Subject: 988	LVAGTTGSGKSVGVNAMILSMLYKAQPEDVRFIMIDPKMLELSYSGIPLHLLTEVVTDMK	1047				
50	Query: 736	LAANLNCVNMCKRYRLMSFMGVRNLAFGNQKIAEAAARGEKIGNPFSLTPODPEP--	793				
		ANAL NCVNMCKRYRLMS +GVNLAG+N+KIAEA I +F+ D +					
	Subject: 1048	DAANLNCVNMCKRYRLMSALGVNLAGNYEKIAEADRMRRFPIDPYMKPGDSMDAQH	1107				
55	Query: 794	--LEKLPIVVVVVDEPADLMMTAGKKIEELIARLAQKARAAGIHLILATQRPVSDVITGL	851				
		+L+K +IVV+VDEPADLMMT GKK+ELIARLAQKARAAGIHL+LATQRPVSDVITGL					
	Subject: 1108	PVLKKEPYVVLVDEPADLMMTVGKKVEELIARLAQKARAAGIHLVILATQRPVSDVITGL	1167				
	Query: 852	IKANIPTRIAFQVSSKIDSRITLQDMGAENLLGGQDMFLPLPGTAYPQRVHGAFASDEEV	911				
		IKANIPTRIAF VSSKIDSRITLQ GAE+LLG GDMFL P + P RVHGAF D+EV					
60	Subject: 1168	IKANIPTRIAFTVSSKIDSRITLQDGAESLLGMDMLYSGNPSTLPRVHGAFVDRQEV	1227				
	Query: 912	HRVVEYLKQFGPEPDYVDDILSGGGSEELPGIGRSGDGETDPMYDEAVSVVLKTRKASISG	971				
		H VV+ K G F YVD I S SE G G G E DP++D+AV V +RKASISG					
	Subject: 1228	HAVVQDKARGRFQYVDGITSDESSEGGAG-GFDGAESLPLFDQAVQVTEKRRKASISG	1286				
65	Query: 972	VORALRIGYNRAARLIDQMEAGIVSAPEHNGNRTILVPL	1010				
		VOR RIGYNRAAR+I+QMEA+GIVS HNGNR +L P					
	Subject: 1287	VORQFRIGYNRAARIIQMEA+GIVSEQINGNREVLAP	1325				

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 59

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 497>:

```

5      1  ATGATTATC  AAAGAAACCT  CATCAAAGAA  CTCCTCTTTA  CGCGCGTGGG
    51  CATTTCGCTC  GTCCCTCTGG  CGGTATTGGT  CTCACGACAG  GCAATCAACC
   101  TGCTCGGCGG  TGCGCGCGAC  GGGC...GTGA  TCGCCATCGA  TGCCGTGTGG
   151  GCATTGGTGC  GCTTCTGGGT  C.....

10     .....A  TTGCCATCGG  TTTGTTTTTA  ATTTACCAAA  ACGGCGTGAC
   91  CCTGCTTTTT  GAAGCGGTGG  AAGACGGCAA  AATTCATTTT  TGGCTCGGAC
  101  TGCTGCTAT  GCACATTATC  ATGTTTGTC  TTGCATCAT  CCGTGTGGGC
  1051  GTCCGACGTA  TGCCGACCCA  GCCTTCTGCG  CAGGCGGTGG  GC AAAAGTCT
  1101  GACATTGAAA  GCGCGAAAT  GA

```

15 This corresponds to the amino acid sequence <SEQ ID 498; ORF101>:

```

      1  MIYORNIKE  LSFTAVGIFV  VLLAVLVSTQ  AINLGRAAD  GXVIAIDAVL
    51  ALVGFVW... ..

20     ...IAIGLFL  IYQNGLTLLF  EAVEDGKIHF  WLGLLPHMII  MFVLALILLR
   301  VRSMPSQPFW  QAVGKSLTLK  GKG*

```

Further work revealed the complete nucleotide sequence <SEQ ID 499>:

```

      1  ATGATTATC  AAAGAAACCT  CATCAAAGAA  CTCCTCTTTA  CGCGCGTGGG
    51  CATTTCGCTC  GTCCCTCTGG  CGGTATTGGT  CTCACGACAG  GCAATCAACC
   101  TGCTCGGCGG  TGCGCGCGAC  GGGC...GTGA  TCGCCATCGA  TGCCGTGTGG
   151  GCATTGGTGC  GCTTCTGGGT  C.....

25     .....A  TTGCCATCGG  TTTGTTTTTA  ATTTACCAAA  ACGGCGTGAC
   91  CCTGCTTTTT  GAAGCGGTGG  AAGACGGCAA  AATTCATTTT  TGGCTCGGAC
  101  TGCTGCTAT  GCACATTATC  ATGTTTGTC  TTGCATCAT  CCGTGTGGGC
  1051  GTCCGACGTA  TGCCGACCCA  GCCTTCTGCG  CAGGCGGTGG  GC AAAAGTCT
  1101  GACATTGAAA  GCGCGAAAT  GA

35     .....A  TTGCCATCGG  TTTGTTTTTA  ATTTACCAAA  ACGGCGTGAC
   91  CCTGCTTTTT  GAAGCGGTGG  AAGACGGCAA  AATTCATTTT  TGGCTCGGAC
  101  TGCTGCTAT  GCACATTATC  ATGTTTGTC  TTGCATCAT  CCGTGTGGGC
  1051  GTCCGACGTA  TGCCGACCCA  GCCTTCTGCG  CAGGCGGTGG  GC AAAAGTCT
  1101  GACATTGAAA  GCGCGAAAT  GA

45     .....A  TTGCCATCGG  TTTGTTTTTA  ATTTACCAAA  ACGGCGTGAC
   91  CCTGCTTTTT  GAAGCGGTGG  AAGACGGCAA  AATTCATTTT  TGGCTCGGAC
  101  TGCTGCTAT  GCACATTATC  ATGTTTGTC  TTGCATCAT  CCGTGTGGGC
  1051  GTCCGACGTA  TGCCGACCCA  GCCTTCTGCG  CAGGCGGTGG  GC AAAAGTCT
  1101  GACATTGAAA  GCGCGAAAT  GA

```

45 This corresponds to the amino acid sequence <SEQ ID 500; ORF101-1>:

```

      1  MIYORNIKE  LSFTAVGIFV  VLLAVLVSTQ  AINLGRAAD  GRVAIDAVLA
    51  LVGFVWIGMT  PLLLVLTAFI  STLTVLTRYW  RDEMSVWLVS  CGLALKQVIR
   101  PVMQFAVPFA  VLVAVMQLMW  IPWAEILRSRE  YAEILKQKQE  LSLVEAGEFN
   151  SLGKRNGRVY  FVETFDTESS  IMKNLFLREQ  DRNGGDNILF  AREGNFSLMD
   201  NKRTLELRHG  YRGSTPAGRA  DYNQVSFKRL  NLIISTPKL  IDPVSHRRTI
   251  PFOILGSEN  PQHQAELMR  LSLTVSLVLL  CLLAVPLSYF  NERSGHTYNI
   301  LIAIGLFLY  QNGLTLLFEA  VEDGKIHFVL  GLLPWHIIMF  AVAILILVR
   351  SMPSPFWQA  VGKSLTLKGG  K*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF101 shows 91.2% identity over a 57aa overlap and 95.7% identity over a 69aa overlap with an ORF (ORF101a) from strain A of *N. meningitidis*:

5	orf101.pep	MIYQRNLKELSF	10	20	30	40	50
	orf101a	MIYQRNLKELSF	10	20	30	40	50
10	orf101.pep	90	100	110		
	orf101a	LTVSVLLLC	280	290	300	310	320
15	orf101.pep	120	130	140	150	
	orf101a	LMFHIMFVIA	340	350	360	370	

The complete length ORF101a nucleotide sequence <SEQ ID 501> is:

1	ATGATTATC	AAAGAAACCT	CATCAAGAA	CTCTCTTTA	CGCCGTCGG
51	CATTTTCGC	GTCCCTCTGG	CGGTATTGG	CTCCACGAC	GCAATCAAC
101	TGCTCGGCC	TGCCGCGAC	NGCGTNTCG	CCATCGATC	CGTGTGGCA
151	TTGGTCGGC	TCTGGGTNN	NGGNATGAC	CGCTTTTGC	TNGTGTGAC
201	CGCATTATC	AGTACGTGA	CGGTGTGAC	CGCTACTGG	CNGACAGCG
251	AAATGTCGT	CTGGNTATC	TGCGGATTG	CATTGAAAC	ATGGATACG
301	CGSGTGATC	AGTTTGGCT	GCGGTTGCG	GTTTGGTGT	CGCTCATCA
351	GCCTTGGGT	ATACCGTGG	CAGACGTAC	CAGCCGCGA	TACGCTGAA
401	TCTCAAGCA	GAAAGTGG	TTGTCTTTC	TGGAGCAGG	CGGGTCAAC
451	AGTTTGGCA	AGCGCAACG	CAGGTTTAT	TTTGTGAAA	CCTCTGATC
501	CGAATCGGC	ATCATGAAA	ACCTGTTCT	GCGGCAACG	GACAAAAAC
551	GCGGCGACA	CATCATCTT	NCCAAAGAA	GTAACCTTC	GCTGAACAC
601	AACAAACGA	CGCTCGAAT	GCGCCACGC	TACGTTTAC	GCGGACGCC
651	CGGACGCGC	GACTACAAT	AGGTTTCTT	CNAAAACTC	AACCTGATT
701	TCAGCACCA	GCCCAACTC	ATCGACCCG	TTTCCACCG	CCGTACNAT
751	CNACNGCCC	AACGTATTG	CAGCAGCAAC	CGGCACATC	ANGCGGAAT
801	GATGTGGCG	ATCTCGTGA	CGCTCAGCT	CCTCTTACT	TGCTGCTTG
851	CGTGCCGCT	TTCTATTTC	AACCCGCGA	GCGACATAC	CTACAATAT
901	TTGANTGCC	TGGGTTGTT	TTTAATTAC	CAAAACGGG	TGACCTGCT
951	TTTGAAGCC	GTGGAGACG	GCAAAATCA	TTTTGGCTC	GGACTGCTG
1001	CTATGCACAT	CATCATGTC	GTCATCGCA	TGTACTTCT	GCSCGTCGC
1051	ASCAAGCCA	GCCAGCCTT	CTGGCAGCG	GTTGGCAAA	GTCTGACAT
1101	GAAAGCGGA	AAATGA			

45 This encodes a protein having amino acid sequence <SEQ ID 502>:

1	MIYQRNLKE	LSFTAVGIFV	VLLAVLVSTQ	AINLIGXAAD	XRXAIDAVLA
51	LVGFWXXMT	PLLLVLTAFI	STLTVLTRYW	RDESMVWXS	CGLALKQWIR
101	PVMQFAPVFA	VLVAVMQLWV	IPWAELSRE	YAEILKQKE	LSLVEAGGFN
151	SLGKRNGRVY	FVETFDTESS	IMKNLFLREQ	DRNGGDNIIIF	XKESNFSLND
201	NKRTLRLRHG	YRYSPTPGRA	DYNQVSPFKL	NLIISTTPKL	IDPVSHRRTX
251	PTAQLIGSSN	POHXAELMWR	ISLTVSVLL	CLLAVPLSYF	NPRSGHTYNI
301	LXAIGLFLYI	QNGSLTLFEA	VEDGRIHFWL	GLLPMHIIME	VIAIVLLVRV
351	SMPSQFWQA	VGKSLTLKGG	K*		

ORF101a and ORF101-1 show 95.4% identity in 371 aa overlap:

55	orf101a.pep	MIYQRNLKELSF	60
	orf101-1	MIYQRNLKELSF	60
60	orf101a.pep	PLLLVLTA	120
	orf101-1	PLLLVLTA	120

5	orf101a.pep	IPWAEILRSREYAEILKQKQELSLVEAGGFNSLGKRNGRVYFVETFDTESGIMKNLFLREQ	180
	orf101-1	IPWAEILRSREYAEILKQKQELSLVEAGGFNSLGKRNGRVYFVETFDTESGIMKNLFLREQ	180
10	orf101a.pep	DRNGGDNII FXKESNFSNDNKRTELRHGYRYSGTPEGRADYNQVSFXKLNLIISTTPKL	240
	orf101-1	DRNGGDNII FFAKEGNSFNNDNKRTELRHGYRYSGTPEGRADYNQVSFQKLNLIISTTPKL	240
15	orf101a.pep	IDPVSHRRITPTAQLIGSSNPQHAEIMWRISLTVSVLLCLLAVPLSYFNPRSGHTYNI	300
	orf101-1	IDPVSHRRITPTAQLIGSSNPQHAEIMWRISLTVSVLLCLLAVPLSYFNPRSGHTYNI	300
20	orf101a.pep	LXAIIGFLIYQNGLTLLFEAVEDGKIHFWLGLLPMHIIMFVIAIVLLRVRSMSPQFWQA	360
	orf101-1	LXAIIGFLIYQNGLTLLFEAVEDGKIHFWLGLLPMHIIMFAVALILLRVRSMSPQFWQA	360
	orf101a.pep	VGKSLTLKGGK	371
	orf101-1	VGKSLTLKGGK	371

Homology with a predicted ORF from *N.gonorrhoeae*

ORF101 shows 96.5 % identity in 57aa overlap at the N-terminal domain and 95.1% identity in 61aa overlap at the C-terminal domain, respectively, with a predicted ORF (ORF101ng) from *N.gonorrhoeae*:

25	orf101.pep	MIYQRNLIELSFTAVGIFVLLAVLVSTQAINLLGRAADGXVIAIDAVLALVGFWV	57
	orf101ng	MIYQRNLIELSFTAVGIFVLLAVLVSTQAINLLGRAADGRV-AIDAVLALVGFWVIGM	59
30		//	
	orf101.pep	IAIGLFLIYQNGLTLLFEAVEDGKIHFWLG	333
35	orf101ng	SLTVSVLLCLLAVPLSYFNPRSGHTYNI LIAIGLFLIYQNGLTLLFEAVEDGKIHFWLG	331
	orf101.pep	LLPMHIIMFVALILLRVRSMSPQFWQAVGKSLTLKGGK	373
	orf101ng	LLPMHIIMFVIAIVLLRVRSMSPQFWQAVG	362

The ORF101ng nucleotide sequence <SEQ ID 503> is predicted to encode a protein having partial

amino acid sequence <SEQ ID 504>:

45	1	MIYQRNLIELSFTAVGIFVLLAVLVSTQAINLLGRAADGXVIAIDAVLALVGFWV
	51	LVGEWVIGMTPLLLVLTAFISTLTVLTRYVRSSEMSVWLSGLALKQWIR
	101	PVHQEAVPEA ILIAVMOLMVIPWAEILRSREYAEILKQKQELSLVEAGGFNS
	151	NLGRNGRVYFVFETPTESGIMKNLFLREQDRNGGDNIIIFAKEGNSFNNDNKRTE
	201	NKRTLELRHGYRYSGTPEGRADYNQVSFXKLNLIISTTPKLIDPVSHRRITPTAQL
	251	IGSSNPQHAEIMWRISLTVSVLLCLLAVPLSYFNPRSGHTYNI
	301	LXAIIGFLIYQNGLTLLFEAVEDGKIHFWLG
	351	SMSPQFWQAVG...

Further work revealed the complete nucleotide sequence <SEQ ID 505>:

50	1	ATGATTATCATCAAGAAACCTCATCAAGAACTCTCTCTTTACGCGCGTGGG
	51	CATTITTCGTCTCTCTCTGGTGGTGTGGTGTCCACGCGAGCGATCAAC
	101	TGCTTGGCCGCGACGCTGACGGCGGTGTGCGCATCGATGATGCGTGTGGCC
	151	TTAGTCGGCTCTCTGGTGTGATCGGTATGACCGCTTTTGTGTTGAC
55	201	CGCATTCATCAGCAGCGTGAACGATTGACCGCTACTGGCGCGACAGCG
	251	AAATGTCGGTCTGGCTATCCGTGGATGGCGTTGAACATGTGATACGC
	301	CCGCTCATGCAGTTTGGCGTGGCTTTGGCCATCTCTGATGCGTCATCGAA
	351	GCTTTGGGTGATACCGTGGGACAGAGCTGGCGAGCCGCGAATATGCGGAAA
	401	TTTGTGACGACGACGCGAATCTGTCTTGGTGGAAAGCCGCGATTCAT
	451	AACTTGGGACAGCGCAACGCGAGTTCATCTGGAGAACCTTTGACAC
60	501	CGATCCGCGCATCATCAAGAAACCTGTCTCTGGCGACACGACAAAACG
	551	gcgcgacacaCATCATCTTCGCCaagaAGtaactTctcgtcgaggac

5 601 AACAAAcgca cgctcgaaATT GCGCCACGGC TACCGTTACA GCGGcagcgC
 651 CGGacGCGCc gactaCAATC AGGTTtcctt cCAAAAcctc aacctgAtta
 701 TCAGCACCAC GCCCAAcTT ATGccacCG TTTCCACCG CGCACCATT
 751 tcgacCGCCc AAcTGATGG CAGCAGCAAT CGGCAACATC AGGCAGAATT
 801 gATGTGGCGC ATCTCGCTGA CCGTCAGCGT CCTCCTGCTC TGCCTACTCG
 851 CCGTGCCGCT TTCTATTTC AACCCGGGCA GCGGACATAC CTACATATATC
 901 TTGATTGCCA TCGGTTTGT TTTAATTTC CAAAACGGGC TGACCTGTCT
 951 TTTTGAAGCC GTGGAGAGCG GCAAAATCCA TTTTGTGCTC GGACTGCTGC
 1001 CTATGCACAT CATATGTC GTATCGCAA TCGTACTCT CGCGCTCCGC
 1051 AGTATGCCCA GCCAGCCCTT CTGGCAGCG GTTGGCAAAA GTCTGACATT
 1101 GAAAGcgga AATGA

This corresponds to the amino acid sequence <SEQ ID 506; ORF10Ing-1>:

1 MIYQRNLKE LSFTAVGIFV VLLAVLSTQ AINLLGRAAD GRVAIDAVLA
 51 LVGFVVGMT PLLVLTAFI STLTLTRYW RDSEMSVWLS CGLALKQWIR
 101 FVMQFAVPPA ILIAVLMQWV IPWAEIERSRE YAEILKQKQE LSLVEAGEFN
 151 NLGKRNGRVY FVETFDTEG IMKNFLRQEQ DKNGGDNII F AKEGNFSLKD
 201 NKRTLRLRHG YRYSGTPGRA DYNQVFSQKL NLIISTTPKL IDPVSHRRTI
 251 STAQLIGSSN PQHQAEIEMW RLSLTVSVLL CLIAVPLSYF NFRSGHTYNI
 301 LIAIGLFLIY QNGLTLLFEA VEDGKIHFWL GLLPMHIMF VIAIIVLRVR
 351 SMPSPQFWQA VGKSLTLKGG K*

ORF10Ing-1 and ORF101-1 show 97.6% identity in 371 aa overlap:

		10	20	30	40	50	60
orf101-1.pep	MIYQRNLKE	LSFTAVGIFV	VLLAVLSTQ	AINLLGRAAD	GRVAIDAVLA	LVGFVVGMT	
orf10Ing-1	MIYQRNLKE	LSFTAVGIFV	VLLAVLSTQ	AINLLGRAAD	GRVAIDAVLA	LVGFVVGMT	
		10	20	30	40	50	60
orf101-1.pep		70	80	90	100	110	120
orf10Ing-1		70	80	90	100	110	120
		130	140	150	160	170	180
orf101-1.pep	IPWAEIERSRE	YAEILKQKQE	LSLVEAGEFN	SLGKRNGRVY	FVETFDTEG	IMKNFLRQEQ	
orf10Ing-1	IPWAEIERSRE	YAEILKQKQE	LSLVEAGEFN	SLGKRNGRVY	FVETFDTEG	IMKNFLRQEQ	
		130	140	150	160	170	180
orf101-1.pep		190	200	210	220	230	240
orf10Ing-1		190	200	210	220	230	240
		250	260	270	280	290	300
orf101-1.pep	IDPVSHRRTI	PTAQLIGSSN	PQHQAEIEMW	RLSLTVSVLL	CLIAVPLSYF	NFRSGHTYNI	
orf10Ing-1	IDPVSHRRTI	PTAQLIGSSN	PQHQAEIEMW	RLSLTVSVLL	CLIAVPLSYF	NFRSGHTYNI	
		250	260	270	280	290	300
orf101-1.pep		310	320	330	340	350	360
orf10Ing-1		310	320	330	340	350	360
		370					
orf101-1.pep	VGKSLTLKGGKX						
orf10Ing-1	VGKSLTLKGGKX						
		370					

Based on this analysis, including the presence of a putative leader sequence (double-underlined) and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is

predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 60

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 507>:

5	1	GGTGTCGGTT	TTATCATATGC	TATGCTGTGGCC	ACCTTGAAGCA	CAGGCAAAACC
	51	ATCATATCA	CGAGAGACAG	TTACGGCTTCT	TAGATGATATC	CAGGCAATGTG
	101	TTGTAATCGC	CGGACACGGT	TTGGATGCAC	GTGATACGCA	TTACACACGTT
	151	ATTCTCAAGT	ATCATTCAA	ATATGATCA	CCCGTATGGG	CACAAAGATGT
10	201	TGCGTGTGCT	CGGGGACAA	ACGATGTGGC	CGCAACAGST	GATGCACATT
	251	CGCCATTCT	ATCATATGCT	GCTGCCAATA	GCTCAAAACA	TACGCGCAAT
	301	AGCGGCAC	ATATCCCTTT	ATTTCGCAAT	GATACAGAGA	ATATTAGAGAG
	351	TATCGTGC	CACAAACAA	ACCTTGCAAT	TACGCGTGA	GCAAGCAGCG
	401	ATTGTGTA				

This corresponds to the amino acid sequence <SEO ID 508: ORF113>:

15 1 ..GGGFINASCA TLTTAKPQYQ AGDLSAFK1R QGNVVIAGHG LDARDTDYTR
51 ILSYHSKIDA FVWGQDVRVV AQONDVAATG DAHSPILNNA AANTSNNNTAN
101 NGTHIPLEAI DTGKLGXVC OONHLDYGR ASRHS*

Computer analysis of this amino acid sequence gave the following results:

Homology with with pspA putative secreted protein of *N.meningitidis* (accession AF030941)

20 ORF and *pspA* show 44% aa identity in 179aa overlap:

orf113 GGGFINASCATLTAKPQYQAGDLSAFKIRQGNVVIAGHGLDARDTDYTRILSYHSKIDA 60
GGG INA+ TLT+ P G+L+ F + G VVI G GLD D DYTRILS ++I+A
pspa GGGGLINAASVTLTSGVPVLNNGNLTFGFDVSSGKVVIGGKGLDSDADYTRILSRAAEINA 256

```
orf113  FVWGQDVRVVAGQNDVAATGDAHSPI LXXXXXXXXXXXXXGTHIPLFAIDTGLGMYA 120
        VWG-DV-VV-G-N + G + P AIDT LGMYA
pspa    GVWGKDVKVVS GKNKLD FGDG-----SLAKTASAPSSSDSVTPTVAIDTATLGMYA 307
```

30 orf113 NKITLISTVEQAGIRNQGWFFASAGNVAVNAEGKLVNTGMIAATGENHAVSLHARNVHN 179
+KITLIST A IRN+G+ FA+ G V ++A+GKL N+G I A +++ A+ V N
pspa DKITLISTDNGAVIRNKGRIFAATGGVTLSDAGKLSNNGSIDAA----ETTISAQTVDN 362

Homology with a predicted ORF from *N. gonorrhoeae*

ORF113 shows 86.5% identity in 52aa overlap at the N- terminal part and 94.1% identity in 17aa

35 overlap at the C-terminal part with a predicted ORF (ORF113ng) from *N. gonorrhoeae*:

orf113 GGGFINASCATLTAKPOYOAGDLSAFKIR 30

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100

40 orf113 QGNVVIAGHGLDARDTDYTRILSYHSKIDAPVWGQDVRVVAGQNDVAATGDAHSPILNNA 90

orfl13ng QGNAVIAGHGLDARDTDFTRILVCCQNHLDQYGRTSRHS 263

orf113 IDTGKLGXVCQONHLDQYGRASRHS 135

45 orf113nq DFGGFKIRQGNNAVIAGHGLDARDTDETRILVCOONHLDQYGRTRHS 263

The complete length ORF113ng nucleotide sequence <SEQ ID 509> is predicted to encode a protein having amino acid sequence <SEO ID 510>:

50 1 MNKTLYRVIF NRKRGAVVAV AETTKREGKS CADSGSGSVY VKSVSFIFTH
51 SKAFCFSLG FSLCLALGTV NIAFADGIIT DKAAPKTOQA TILQTNGIIP

-300-

```

101 QVNIQTPTSA GVSVNQYQF DVGNRGAILN NSRSNTOTOL GGWIGQNPWL
151 TRGEARVVVN QINSSHFSQL NGYIEVGGRR AEVVIANFAG IAVNGGGFIN
201 ASRATLTGQ PQYQAGDFSG FKIRQGNVAVI AGHGLDARDT DTRILVCCQO
251 NHLDOYGRTS RHS*
```

- 5 Based on this analysis, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 61

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 511>:

```

101 1 ..TCAACGGGAC ATAGCGAACA AAATTACACT TTGCGCGGAG AAATCACAGC
51 51 CAACATTTTCA CTGGGTTTCAT TTGCTTATGA ATCGCATCGC AAAGCATTA
101 101 GCCATCATGC GCCAGCGCAA GGCACGTAGT TGCCGCAAG CAACGTTAT
151 151 TCGTACCTCT ATACGTCCAA TTCTTTTACC CCATTACCCA CGCAGCAGCT
201 201 ATACATTATC AATCTGTGTA ATAAAGGCTA TCTTTGTGAA ACCGATCCAC
251 251 GCTTTGCCAA CTACCGTCAA TGCTTGGGTA GTGACTATAT GCTGGACAGC
15 301 CTCAACTAG ACCCAACCAA TTTACATAA CGTTTGGGTG ATGGTTATTA
351 351 CGAGCACAGT TTAATCAATG AACAATCGC AGAGCTGACA GGGCATCGTC
401 401 GTTTAGACGG TTATCAAAAC GACGAAGAAC AATTTAAGC CTTAATGGAT
451 451 AATGGCGCGA CTGCGGCAGC TTGCGATGAT CTCAGCGTTG GCATTTCGAT
501 501 AAGTACCGG CAGTACGCGC AACTGACGAG CGATATGTTT TGGTTGGTAC
20 551 AAAAAGAAGT TAAGCTTCTT GATGGCGGCA CACAAACCGT ATTGGTGCCA
601 601 CAGGTTTATG TAGCGCTTAA AATGGCGGAC ATAGACGGTA AAGTGCATT
651 651 GTTGTACAGC AGCAATACAC AAATCAATGT TTCAGGCAGC CTGAAAAATC
701 701 CAGGCACGAT TGCAGGCGC AATGCGCTTA TTATCAATAC CGATACGCTA
751 751 GACATATATC GTGGGCGTAT TCATGCGGCA AAATCAGCGG TTACGGCCAC
25 801 ACAAGCATCT AATAATATTG GCGGATGCTT TTCTGCCGAA CAGACATTAT
851 851 TGCTCAACGC AGGCACACAC ATCAACAGCC AAAGCACCAC CGCCAGCAGT
901 901 CAAATACAC AAGGCAGCAG CACCTACCTA GACCSAATGG CAGGTATTTA
951 951 TATCACAGGC AAGAAAAAAG GTGTTT..
```

This corresponds to the amino acid sequence <SEQ ID 512; ORF115>:

```

301 1 ..STGHSEQNYT LPREITRNIIS LGSFAYESHK KALSHHAFSQ GTFLPQSNIGI
51 51 SLFYTSNSFT PLPSSSLYII NFVNGKYLVE TDFRFRANYRQ WLGS DYMLDS
101 101 LKLDNNLHK RLCGQYIEQR LINEQIAELT GHRRLDGYON DEEQFKALMD
151 151 NGATAARSNN LSVGIALSAE QVAGLQTSIV WLVOKEVKLF DSGTQVLVLF
35 201 QVYVRVKNIG IDGKGALLSG SNTQINVSQS LKNSCTIAGR NALIINTDTL
251 251 DNIGGRINHAQ KSAVTATQDI NNIGGMLSAE CILLINAGNN INSQSTTASS
301 301 QNTQGSSTYL DRMAGIYITG KEKGV..
```

Computer analysis of this amino acid sequence gave the following results:

Homology with the pspA putative secreted protein of *N.meningitidis* (accession number AF030941)

ORF115 and pspA protein show 50% aa identity in 325aa overlap:

```

401 Orf115: 1 STGHSEQNYTLPREITRNIISLGSFAYESHKRALSHHAFSQGTFLPQSNIGISLPYTSNSFT 60
STG+ S Y E++ +I +G AY+ + + P + NGI +T
pspA: 778 STGYSRSPYEPAPPEVS-SIRMGISAYKYAPQQASDIPGTUVVPAVENGIHPTFT----- 831

451 Orf115: 61 PLPSSSLYIINPVNKGYLVE TDFRFRANYRQWLGS DYMLDSLKDNNLHKRLGDDGYEQR 120
LP+SSL+ I P NKGYL+ETDF F +YR+WLGS YML +L+ DPN++HKRLGDDGYEQ+
pspA: 832 -LPNSSLFAIAPNNKGYLIETDPAFTDYRKWLGSYMLAALQDDPNHIIKRLGDDGYEQK 890

501 Orf115: 121 LINEQIAELTGHRRLDGYQNDDEQFKALMDNGATAARSNNLSVGIASAEQVAQLTSDIV 180
L+NEQIA+LTG+RRLDGY NDEEQFKALMDNG T A+ + L+ GIALSAEQVA+LTSDIV
pspA: 891 LVNQEIAKLTGYRRLDGYNDDEQFKALMDNGITIAKEQLTPIGIALSAEQVAQLTSDIV 950

551 Orf115: 181 WLVOKEVKLPDGGTQTVLVPQVYVRVKNIGIDGKGALLSGSNTQINVSGLKN-SGTIAG 239
NL + V LPDG TQTVL P+VYVR + D+G+GALLSGS I SG++N G IAG
pspA: 951 WLENETVTLPDGTQTVLKP+KVYVRARPKDMNGQALLSGSVVDIG-SGAIENRGGLIAG 1009
```

Orf115: 240 RNALITINTDTLDNIGGRHAQSAVTAQDINNIGMGLSAEQTLNLAGXXXXXXXXX 299
 R ALT+N + N+ G + + A DI N G + AE LLL A
 pspA: 1010 REALITNNAQNIKNLQDLQGNIFAAAGSDITNTGS-IGAENALLKASNNIESRSETRS 1068

5 Orf115: 300 XXXXXXXXXYLDRMAGIYITGKEGK 324
 + R+AGIY+TG++ G
 pspA: 1069 NQNEQGSVRNIGRAGVYLTGRQNG 1093

Homology with a predicted ORF from *N.gonorhoeae*

- 10 ORF115 shows 91.9% identity over a 334aa overlap with a predicted ORF (ORF115ng) from *N.gonorhoeae*:

orf115.pep	STGHSEQNYTLPREITRNISLGSFAYESHK	31
orf115ng	NEQTFGEEKVFSENGKLNHYWRARRKGHDGHRQNYTLPEBITRDISLGSFAYESHK	71
orf115.pep	ALSHHAPSQGTLPQSN-----GISLPYTSNFTPLPSSSLYIINPVNKGVLVET	81
orf115ng	ALSRHAPSQGTLPQSNRDNIRTAQSNGLSIPYTPNSFTPLPSSSLYIINPVNKGVLVET	131
orf115.pep	DPRFANYRCWLGSDYMLDSLKLDPNNLHKRLGDGYEQRLINEQIAELTGHRRLDGYQND	141
orf115ng	DPRFANYRCWLGSDYMLGSLKLDPNNLHKRLGDGYEQRLINEQIAELTGHRRLDGYQND	191
orf115.pep	EEQFKALMDNGATAARSMNLSVGIASAEQVAQLTSDIVLVQKEVKLPDGGTQTVLVPQ	201
orf115ng	EEQFKALMDNGATAARSMNLSVGIASAEQVAQLTSDIVLVQKEVKLPDGGTQTVLMPQ	251
orf115.pep	VYVRVKNQDIDGKGALLSGSNQTQINVSGLSKSGTIAGRNALIINTDTLDNIGGRHAQK	261
orf115ng	VYVRVKNQSIDGKGALLSGSNQTQINVSGLSKSGTIAGRNALIINTDTLDNIGGRHAQK	311
orf115.pep	SAVTATQDINNIGMGLSAEQTLNLAGNNINSQSTASSONTQGSSTYLDRMAGIYITGK	321
orf115ng	SAVTATQDINNIGMGLSAEQTLNLAGNNINSQSTAKSQNAQGSSTYLDRMAGIYITGK	371
orf115.pep	EKGV	325
orf115ng	EKGVLAAQAGKIDINIAGQISNSQSDQGTQLQAGRIDNLDTVQTKGYQEIHFDDADNHTIR	431

- 40 An ORF115ng nucleotide sequence <SEQ ID 513> was predicted to encode a protein having amino acid sequence <SEQ ID 514>:

1	MLVQTEKDGL	HNEQTFGEKK	VFSENGKLN	YWRARRKGHD	ETGHRQNYT
51	LPPEITRDIS	LGSFAYESH	KALSRHAPSQ	GTELPQSNRD	NIRTAQSNGI
101	SLPYTFNSFT	PLPGSSLYII	NPANKGYLVE	TDPFRFANYRQ	WLGSDYMLGS
151	LKLDPNNLHK	RLGDGYEQR	LINEQIAELT	GHRRLDGYQN	DEEQFKALMD
201	NGATAARSMN	LSVGIASAE	QAAQLTSDIV	WLVQKEVKLP	DGGTQTVLVPQ
251	QYVRVKNQD	IDGKGALLSG	SNQTQINVS	GSLKSGTIAG	RNALIINTDTL
301	DNIGGRHAQ	KSAVTATQDI	NNIGGLLSAE	QTLNLAGNN	INSQSTAKSS
351	QNAQGSSTYL	DRMAGIYITG	KEKGVLAQA	GKDINIAGQ	ISNSQDGGT
401	RLQAGRIDNL	DTVQTKGYQ	EIFDADNHTI	RGSTNEVGS	IQTKGDVTL
451	SGNNLAKAA	EVGSAGKTLA	VYAKNDITIS	SGIHAGQVDD	ASKHTGRSGG
501	GNKLVIDTKA	QSHHETAQSS	TPEGQVVLQ	AGNDANILGS	NVISDNTRI
551	QAGNHVRIGT	TQTQSQSEY	HTQKSGSLMS	AGIGFTIGSK	TNTQENQSQS
601	NEHTGSTVGS	LKGDITIVAS	KHYEQTGSNV	SSPEGNNLIS	TQSMDIGAAQ
651	NQLNSKITQT	YEQKGLTVAF	SSPVTDLAQ	ALAVAHKA	QFDKAKTTAL
701	MPWRLEMQVG	RLFKQAKAPK	K*		

Further work revealed the following partial gonococcal DNA sequence <SEQ ID 515>:

- | | | | | | |
|-----|------------|------------|------------|-------------|-------------|
| 1 | TTGCTIGTGC | AAACAGAAAA | AGACGGTTTG | CATAACGAGC | AAACCTTTGG |
| 51 | CGAGAGAGAA | GTCTTCAGC | AAATGGTTAA | GTGCACAC | TACTGGCGTG |
| 101 | CGCGTCGTAA | AGGACATGAT | GAACACGGGC | ATCGTAGACA | AAATATATCT |
| 151 | TTGCCGAGAG | AAATCACACG | CGACATTTC | CTGGGTTTCAT | TGCTCATATGA |
| 201 | ATCGCATAGC | AAAGCATTAA | CGCGTCATGC | GCCACGCCAA | GGCAGTGAAT |
| 251 | TGCCACAAAG | TAACCGGGAT | AATATCCGTA | CTGCGAAGAG | CAACGGTATT |

301	TGCTACCOCT	ATACGCCCAA	TCTTTTACC	CCATTACCCG	GCAGCAGCTT
351	ATACATTATC	AATCCTGCCA	ATAAAGGCTA	TCTTGTGAA	ACCGATCCAC
401	GCTTTGCCAA	CTACCGTCAA	TGGTGGGTA	GTGACTATAT	GCTGGGCGAG
451	CTCAAACTAG	ACCCAAACAA	TTTACATAAA	CGTTTGGGTG	ATGGTTATTA
501	CGAGCAACGT	TTAATCAATG	AACAAATCGC	AGAGCTGACA	GGCATCGTC
551	GTTTAGACGG	TTATCAAAAC	GACGAAGAAC	AATTTAAAGC	CTTAATGGAT
601	AATGGCGCGA	CTGGGCGACG	TTCGATGAAT	CTCAGCGTTC	GCAATTGCATT
651	AGTGGCGGAG	CAAGCAGCGC	AACCTGACCG	CGATTTTGT	TGGTGGGTAC
701	AAAAAGAGAT	TAAATCTTCT	GATGGCGGCA	CACAAACCGT	ATTGATGGAG
751	CAGGTTTATG	TAGCGGTTAA	AAATGGCGCG	ATAGACGGTA	AAAGTGCATT
801	GTGTGCAGCG	AGCAATACAC	AAATCAATGT	TTCAGGCAGC	CTGAAAAACT
851	CAGGCAOGAT	TGCAGGGCGC	AATGCGCTTA	TTATCAATAC	CGATACGCTA
901	GACAAATATCG	GTGGGCGTAT	TCATGCGCAA	AAATCAGCGG	TTACGGCCAC
951	ACAAGACATC	AATAATATTG	GCGGCATTCT	TTCTGCCGAA	CAGACATTAT
1001	TGCTCAATGCG	GGGTAAACAC	ATCAACAACC	AAAGCAGCGC	CAGAGCAGT
1051	CAAAATGCGAC	AAGGTAGCAG	CACCTACCTA	GACCGAATGG	CAGGTATTTA
1101	TATCACAGCG	AAAGAAAAAG	GTGTTTACG	AGCGCAGGCA	GGCAAAAGCA
1151	TCAACATCAT	TGCCGGTCAA	ATCAGCAATC	AATCAGATCA	AGGGCAAAAC
1201	CGCGTGCAGG	CAGGACGCGA	CATTAACTCG	GATACGGTAC	AAACCGGCAA
1251	ATATCAAGAA	ATCCATTTTG	ATGCGCGATA	CCATACCATC	CGAGGTTCAA
1301	CGAACGAGAT	CGCGACGAGC	ATTCAACAAA	AGGCGGATGT	TACCClaTG
1351	TCAAGGAATA	ATCTCAATGC	CAAACTGGCC	GAACTCGGCA	GCGCAAAAGG
1401	CACACTTGGC	GTGTATGCTA	AAAAAGCAT	TACTATCAGC	TACGCGATCT
1451	ATGCCGCGCA	AGTTGATGAT	CGGTCCAAC	ATACAGGCGC	AAAGCGCGCG
1501	GGTAATAAAT	TAGTCATTAC	CGATAAAGCC	CAAAGTCATC	ACGAATACTGC
1551	TCAAGCGAGC	ACCTTTGAAG	GCAAGCAAGT	TGTATTGACG	GCAGGAAACG
1601	ATGCCAACAT	CCTTGGCAGT	AATGTTATTT	CCGATAATGG	CACCGCGATT
1651	CAAGCAGGCA	ATCATGTTTG	CATTGGTACA	ACCCAAAGCT	AAAGCCAAAG
1701	CGAAACCTAT	CATCAACACC	AAAAATCAGG	ATTGATGAGT	CGAGGTATCG
1751	GCTTCACTAT	TGCCAGCAGG	ACAAACACAC	AAGAAACCA	ATCCCAAAAG
1801	AAAGAACATA	CAGCGAGTAC	CGTAGGCGAG	CTGAAAGGCG	ATACACCAT
1851	TGTTGCADGC	AAACACTACG	AACAACAACG	CGACAAGT	TCCAGCCGCT
1901	AGGCGAACAA	CCTTATCAGC	ACGCAAAATA	TGGATATTGG	CGCAGCACAA
1951	AAACATTTAT	ACAGCAACAG	CACCCAAAGC	TACGACAAA	AAAGCTTAA
2001	GCTGGCATTC	ACTTGGCGCG	TTACAGGATT	GGCAGACAA	CGATATGCTT
2051	TAGCACACAA	EVGSAKAAAC	AAGTGGGACA	AGCAAAAC	GACCGGCTTA
2101	ATGCCATGCG	GGCTGCGAAT	GCAGTGTGCG	AGGCCTATCA	AACAGGCAAA
2151	GGCGACAAA	ACTTAG			

This corresponds to the amino acid sequence <SEQ ID 516; ORF115ng-1>:

40	1	LLVQTEKDGL	HNEQTFFGKK	VFSENGLHN	YWRARRKGDH	ETGHREQNYT
	51	LPPEITRDIS	LGSFAYESH	KALSRHAPSQ	GTLPQSNRD	NIRTAKSNGI
	101	SLPYTPNSFT	PLFGSSLYII	NPANKGYLVE	TDPRFANYRQ	WLGSDYMLGS
	151	LKLDPNNLHK	RLGDGYVEQR	LINEQIAELT	GHRRLDGYQN	DEEQFKALMD
	201	NGATAARSMN	LSVGIALSAE	QAQLTSDIV	WLVOKEVKLP	DGGTQTVLMP
	251	QVYVRVKNNG	IDHGALLSG	SNQTNVSGS	LRNSGTIAGR	NALLINTDTL
	301	DNIGGRHQAQ	KSAAVTATQI	NIGGILSAE	QTLNLNAGNN	INNOSTAKSS
	351	QAQGSSTYL	DRAGAGIYTG	KEKYLAAQA	GKDNLILAQ	ISNOSDQGT
	401	RLQAGRDINL	DTVQTSKYQE	THPDADHHT	RSNTNEVSS	IQTKSDVILL
	451	SGNLLNKAQA	EVGSAKGYLA	YYAKNDITIS	SGIHAGQVDD	ASKHTGRSGG
	501	GNKLVIITDKA	QSHHETAQSS	TFFEGQVVLQ	AGNDANILGS	NVISDNGTRI
	551	QAGNHVRIGT	TQTOSQSEYI	ROTQKSLMS	AGIGFTIGSK	TNTQENQOSQ
	601	NEHTGSTVGS	LKGDITTVIS	KHYEQTGSNV	SPPEGNLLIS	TQSMIDGAAQ
	651	NQLNSKTQY	YEQKGLTVAF	SSPVTDLAQQ	ATAVAHKAAN	KSKDAKTTAL
	701	MPWRLPMQVG	RFIKQAKAHK	T*		

55 This gonococcal protein (ORF115ng-1) shows 91.9% identity with ORF115 over 334aa:

	20	30	40	50	60	70
orf115ng-1.p	NEQTFFGKKVFSENGLHN	YWRARRKGDH	ETGHREQNYT	LPPEITRDIS	LGSFAYESHK	
orf115				STGHSEQNYTL	PREITRNISLGS	FAYESHK
60				10	20	30
	80	90	100	110	120	130
orf115ng-1.p	ALSRHAPSQGTLPQSNRD	NIRTAKSNGIS	LPYTPNSFT	PLFGSSLYII	NPANKGYLVE	
65	orf115	ALSHHAPSQGTLPQSN	-----	GISLPYTPNSFT	PLFGSSLYII	NPANKGYLVE
	40	50	60	70	80	

		140	150	160	170	180	190
	orf115ng-1.p	DFRFANYROWLGS	DMYLGSLKLD	PNNLHKRLG	DGYEQR	LINEQIAELT	GHRRLDG
	orf115	DFRFANYROWLGS	DMYLGSLKLD	PNNLHKRLG	DGYEQR	LINEQIAELT	GHRRLDG
5		90	100	110	120	130	140
		200	210	220	230	240	250
	orf115ng-1.p	EEQFKALMDNGATA	AARSMNLSV	GIASAEQA	QLTSDI	VWL	VQKSVKLP
	orf115	EEQFKALMDNGATA	AARSMNLSV	GIASAEQA	QLTSDI	VWL	VQKSVKLP
10		150	160	170	180	190	200
		260	270	280	290	300	310
	orf115ng-1.p	VYVRVKNGGIDGK	GALLSGSNTQ	INVSGSLKNS	SGTIAGRNALI	INTDTL	DNIGGRIHAQK
	orf115	VYVRVKNGGIDGK	GALLSGSNTQ	INVSGSLKNS	SGTIAGRNALI	INTDTL	DNIGGRIHAQK
15		210	220	230	240	250	260
		320	330	340	350	360	370
	orf115ng-1.p	SAVTATQDINNIG	GILSAEQTL	LLNAGNNIN	NQSTAKSS	QNAQSSSY	LDRMAGIYT
	orf115	SAVTATQDINNIG	GILSAEQTL	LLNAGNNIN	NQSTAKSS	QNAQSSSY	LDRMAGIYT
20		270	280	290	300	310	320
		380	390	400	410	420	430
	orf115ng-1.p	EKGVLAAQAGK	DINIAGQISN	QSDGQTR	LQAGR	DINLDTV	QTKYQIHF
	orf115	EKGVLAAQAGK	DINIAGQISN	QSDGQTR	LQAGR	DINLDTV	QTKYQIHF

In addition, it shows homology with a secreted *N.meningitidis* protein in the database:

30	gi 2623258 (AF030941) putative secreted protein [Neisseria meningitidis] Length = 2273 Score = 604 bits (1541), Expect = e-172 Identities = 325/678 (47%), Positives = 449/678 (65%), Gaps = 22/678 (3%)
35	Query: 1 L1VQTEKDLHNEQTFGEKKVFSGENGLNHYWRARRKHDETHREQNYTLFEEITRDIS 60 L4V T + L N + T G K + ++ G L H V R + K G D T G + Y E + I Sbjct: 739 L1VGTFSALDNDETGLTKTI-TDKGDLHRYHRRHKKGRDSTGYRSFPYEPFVVS-SIR 796
40	Query: 61 LGSFAYESHKALSRHAFSCGTLPQSNRDNIRTAKSNGISLPYTFNSFTPLPGSSLYII 120 +G AY+ + AP Q +++P + + NGI +T LP SSL+ I Sbjct: 797 MGLSAYKGY-----APQASDIPGTV---VPVVAENGIFHTT-----LPNSSLFAI 840
45	Query: 121 NPANKGYLVETDFRFANYROWLGSYMLGSLKLDPNLHKRLGDDGYEQRLINEQIAELT 180 P NKGYL+ETDF F +YR+WLSG YML +L+ DEN++HKRLGDDGYEQ+L+NEQIA+L I Sbjct: 841 APNNKGYLTETDFAFTDYRWKLSGMYLALAQDQPNHILHKRLGDDGYEQKLVNDEQIAKLT 900
50	Query: 181 GHRRLDGYQNDEEQFKALMDNGATAARSMNLSVGIASAEQAQLTSDIWLVLQKSVKLP 240 G+RLDGY NDEEQFKALMDNG T A+ + L+ GIALSAEQ A+LTSDIWL + V LF Sbjct: 901 GYRRLDGYTNDEEQFKALMDNGITAKELQLTGIALSAEQVARTLSDIWLWENETVFLF 960
55	Query: 241 DGTQTVLMPQVYVRVKNGGIDGKALLSGSNTQINVSGSLKNS-SGTIAGRNALIINTDT 299 DG TQTVL P+VYVR + ++G+GALLSGS I SG++N G IAGR ALI+N Sbjct: 961 DGTQTVLKPQVYVRARPKDMNGQALLSGSVVDIG-SGAIENRGLGIAGRALIINQAN 1019
60	Query: 300 LDNIGGRIHAQKSAVTATQDINNIGGILSAEQTLNAGNNINNQSTAKSSQNAQSSSY 359 +N+G + + A DI N G I AE LLL A NNI ++S +S+QN QGS Sbjct: 1020 IKNLQSDLGQKNIFAAGSDITNTGSI-GAENALLKASNNIESRSETRSGNQSQSVRN 1078
65	Query: 360 LDRMAGIYITGKGVLAQAGKIDINIAGQISNQSOGQTRLOAGRDNILDTVQTKYQ 419 +R+AGIY+TGK+G + AG +I + A +++NQS+ GQT L AG DI DT + Q Sbjct: 1079 IGRVAGIYITGRCNGSVLLDAGNNIVUTASELTNQSOGQTVLNAGGDIRSDITGTSRNG 1138
	Query: 420 EIHFADNHTIRGSTNEVSGSIQTKGVDTLISGNNLAKAAVSGAKGTILAVYAKNDITI 479 FD+DN+ IR NEVGS+I+T+G++L + ++ +AAEVGS +G L + A DI + Sbjct: 1139 NTIFSDNYVIRKEQNEVGSITRGNLSNAGKDIRIRAAAEVGSQGRILKLAAGRIDKV 1198
	Query: 480 SSGIHAGQVDDASKITGRSGGNGKLVITDKAQSHHETAQSSTFEKGQVIVLQAGNDANILG 539 +G + +DA K+TGRSGGG K +T ++ + A ST +GK++L +G D + Sbjct: 1199 EAGKHTETEDAKYTGSGGGIKQKMRHLKNQNGQAVSGTLDGKEILIVSGSDITVTG 1258

-304-

Query: 540 SNVISDNGTRIAGNHNVRIGTTQTQSQSEYTHQTKSGLM-SAGIGFTIGSKNTNQENQS 598
 SN-I+DN T + A N++ + +T+S+S ++ +KSGLM S GIGFT GSK +TQ N+S
 Sbjct: 1259 SNIIADNHTILSAKNINIVLKAETRSRSAE MNKESGLMGSGGIGFTAGSKKDTQTNRS 1318

5 Query: 599 QSNEHTGSTVGLSGDITIVASKHIEQTGNSVSPGNNLISTQSMIDAGAAQNLSKTT 658
 ++ HT S VGLS G+T I A KHY QTGS +SSP+G+ IS+ + I AAN+ + ++
 Sbjct: 1319 ETVSHTESVVGSLNGNTLISAGKHYYTQTGSTISSPQGDVGISGKISIDAQNNRYSQESK 1378

10 Query: 659 CTYEQKGLTVAFSSPVT 676
 Q YEQK+TVA S PV +
 Sbjct: 1379 CVYEQKGVTVAFSSPVT 1396

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 62

15 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 517>:

1 ..TCAGGGAATA ACCTCAATGC CAAAGCTGCC GAAGTCAGCA GCGCAACGG
 51 TACACTCGCT GTGTCTGCCA ATAATGACAT CAACATCAGC GCAGGCATCA
 101 ACACGACCCA TGTGTATGAT GCGTCCAAC ACACAGGCAG AAGCGTGGT
 151 GGCATTAAT TAGTCATATC CGATAAAGCC CAAAGTCATC ACGAAACCGC
 201 CCAAGCAGC ACCTTTGAAG GCAAGCAAGT TGTATTGCG GCAGGAAACG
 251 ATGCCAATC CTTTGGCAGC AATGTTATT CCGATAATGG CACCCAGATT
 301 CAAGCAGGCA ATCATGTTGC CATTGCTACA ACCCAAACTC AAAGCCAAAG
 351 CGAACCTAT CATCAAAACC AGAAATCAGG ATTGATGAGT GCAGGTATCG
 401 GCTTCACAT TGGCAGCAGG ACAAAACACAC AAGAAAACCA ATCCCAAGC
 451 AACGAACATA CAGGCAGTAC CGTAGGCAGC TTGAAGGGCC ATACCACCAT
 501 TGTGTCAGGC AAACACTACG AACAAATCGG CAGTACCGTT TCCAGCCCGG
 551 AAGGCAACAA TACCATCTAT GCCCAAGACA TAGACATTCA AGCGGCACAC
 601 AACAAATTA ACAGTAATAC CACCAAAACC TATGAACAAA AAG. CTAA
 651 GGTGGCATTC AGTTCGCCCG TTACGATTT GGCACACAA ...

30 This corresponds to the amino acid sequence <SEQ ID 518; ORF117>:

1 ..SGNNLNKAA EVSSANGTLA VSANNDINIS AGINTTHVDD ASKHTGRSGG
 51 GNKLIVITDKA QSHHETAGSS TFEQKQVVLQ AGNDANILGS NVISDNGTQI
 101 QAGNHNVRIGT TOTQSQSEY THQTKSGLMS AGIGFTIGSK TTNQENQSQS
 151 NEHTGSTVGS LKGDITIVAG KHIEQIGSTV SPPEGNNTIY AQSIDIQAAH
 351 NKLNSTTQT YEQKLTVAF SSPVTDLAQ ...

Computer analysis of this amino acid sequence gave the following results:

Homology with the pspA putative secreted protein of *N.meningitidis* (accession number AF030941)

ORF117 and pspA protein show 45% aa identity in 224aa overlap:

Orf117: 4 NLNAAAEVSSANGTLAVSANNDINISAGINTTHVDDASKHTGRSGGKNKLIVITDKAQSH 63
 ++ +AAEV S G L ++A DI + AG T +DA K+TGRSGG K +T ++
 pspA: 1173 DRRRAAEVSGEQRGLKLAAGRDIKVEAGKHAETEDALKYTRSGGGIKQKMRHLKQ 1232

Orf117: 64 HETAQSSTFEKQVVLQAGNDANILGSNVISDNGTQIQAGNHNVRIGTTQTQSQSEYTHQ 123
 + A S T +GK+++L +G D + GSN+I+DN T + A N++ + +T+S+S ++
 pspA: 1233 NGQAVSGTLDGKEIILVSGRDITVTGNSIIADNHTILSAKNINIVLKAETRSRSAE MNK 1292

Orf117: 124 KSGLM-SAGIGFTIGSKNTNQENQSQSNEHTGSTVGS LKGPFTIVAGKHIEQIGSTVSS 182
 +KSGLM S GIGFT GSK +TQ N+S++ HT S VGLS G+T I AGKHY Q GST+S
 pspA: 1293 EKSGLMGSGGIGFTAGSKKDTQNRSETVSHTESVVGSLNGNTLISAGKHYYTQTGSTISS 1352

Orf117: 183 PEGNNTIYACSIDIQAHNKLNSTTQTYEQKLTVAFFSSPVD 226
 P+G+ I + I I AA N+ + + Q YEQK +TVA S PV +
 pspA: 1353 PQGDVGISGKISIDAQNNRYSQESQVYEQKGVTVAFSSPVT 1396

Homology with a predicted ORF from *N. gonorrhoeae*

ORF117 shows 90% identity over a 230aa overlap with a predicted ORF (ORF117ng) from *N. gonorrhoeae*:

5	orf117.pep	SGNNLNAKAAEVSSANGTLAVSANNDINIS	30
	orf117ng	: : : : :	480
	orf117.pep	AGINTTHVDASKHTRGSGGKNLWITDKAQSHHETAQSSSTFEKGQVVLQAGNDANILGS	90
10	orf117ng	: : : : : : : : : : : : :	540
	orf117.pep	SVIHAGQVDDASKHTGRSGGKNLWITDKAQSHHETAQSSSTFEKGQVVLQAGNDANILGS	
	orf117ng	: : : : : : : : : :	600
15	orf117.pep	NVISDNGTQIQAGNHVRIGTTQTQSQSETYHTQKSGLSAGIGFTIGSKTNTQENQSQS	150
	orf117ng	: : : : : : : : : :	600
	orf117ng	NVISDNGTRIAGNHVRIGTTQTQSQSETYHTQKSGLSAGIGFTIGSKTNTQENQSQS	
	orf117.pep	NEHTGSTVGS LGKDTTIVAGKHVEQIGSTVSSPEGNNITIAQSIDIAAHNKLNSNTTQT	210
	orf117ng	: : : : : : : : : :	660
	orf117ng	NEHTGSTVGS LGKDTTIVASKHVEYQTGSNVSSPEGNNILSTQSMIDIGAAQNQLNSKTTQT	
20	orf117.pep	YEQKXLTAFVSPVTDLAQQ	230
	orf117ng	: : : : : : : : :	720
	orf117ng	YEQKGLTVAFVSPVTDLAQQAIIVAHKAQKQFDKAKTTALMPWRLPMQVGRFLKQAKAPK	

An ORF117ng nucleotide sequence <SEQ ID 519> was predicted to encode a protein having amino acid sequence <SEQ ID 520>:

25	1	..LLVQTEKDKGL HNEQTGEKK VFSENGKLN YWRARRKGHD ETGHREQNYT
	51	LPEZITRDIS LGSFAYESH KALSRHAPSQ GTPLPQSNRD NIRTAKSNGI
	101	SLPYTPNSFT PLPGSSLYII NPANKGYLVE TDPFRANYRQ WLGSQDYLGS
	151	LKLDPNLHK RLGDGYEYQR LINEQIAELT GHRLDGYQN DEEQFKALMD
	201	NGATAARSMN LSVGIALSAE QAAQLTSDIV WLWQKEVKLP DGGTQTVLMP
30	251	QYVVRKNGG IDGKGALLSG SNTQINVS GS LKNSGTIAGR NALIINTDPL
	301	DNIGRIHAQ KSAVTATQDI NNIGGILSAE QTLNLAGNN INNSQAKSS
	351	QNAQSSYTL DRMGITYTG KEKGVLAQA SKDINILAGQ ISNSQGGQT
	401	RLQAGRDIML DTVTQTKYQE THFDADNHTI RSTNVESGS IOTKGVDTLL
35	451	SGNNLNAKAA EVGSAKGTLA VYAKNDITTS SGIHAGQVDD ASKHTRGSGG
	501	GNKLVIDTKA QSHHETAQSS TFEKGQVVLQ AGNDANILGS NVISDNGTRI
	551	QAGNHVRIGT TQTQSQSETY HTQKSGLSMS AGIGFTIGSK TNTQENQSQS
	601	NEHTGSTVGS LGKDTTIVAS KHVEYQTSNV SSPEGNNILS TQSMIDIGAAQ
	651	NQLNSKTTQT YEQKGLTVAF SPSVTDLAQQ AIIVAHKAQK QFDKAKTTAL
	701	MPWRLPMQVG RLKQAKAPK K*

40 Further work revealed the following gonococcal partial DNA sequence <SEQ ID 521>:

	1	TTGCTTGTGC AAACAGAAAA AGACGGTTTG CATAACGAGC AAACCTTTGG
	51	CGAGAAGAAA GTCTTCAGCG AAAATGGTAA GTTGCAACAC TACTGGCGTG
	101	CGCGTCGTAA AGGACATGAT GAACACAGGC ATCGTGAACA AAATTATACT
45	151	TTGCGCGAGG AAATCAACAG CGACATTTCA CTGGGTTTAT TTGCTATGTA
	201	ATGCGATGAC AAACATATA CCCTCATGC GCCACGCCAA GGCCATGATG
	251	TGCCCAAGAG TAACCGGGAT AATTATCGTA CTGCGAAAAC CACGCGATT
	301	TGCTACCTCT ATACGCCCAA TTCTTTTACC CATTATCCCG GCACGAGCTT
	351	ATACATTATC AATCTCGCCA ATAAAGGCTA TCTTGTGTAA ACCGATCCAC
	401	GCTTTGCCAA CTACCGTCAA TGGTTGGGTA GTGACTATAT CTGGGCGAGC
50	451	CTCAAACTAG ACCCAACAAA TTATACATAA CGTTTGGGTG ATGTTTATTA
	501	CGAGCAACGT TTAATCAATG AACAAATCGC AGAGCTGACA GGGCATCGTC
	551	GTTTAGACGG TTATCAAAAC GACGAAGAAC AATTTAAAGC CTTAATGGAT
	601	AATGGCGGGA CTGCGGCACG TCGATGAAT CTCAGCGTTG GCATTGCATT
	651	AAGTGCAGAG CAAGCAGCGC AACTGACCAG CGATATTGTT TGGTTGGTAC
55	701	AAAAAGAGT TAACTTCTCT GATGGCGGCA CACAAACCGT ATTGATGCCA
	751	CAGGTTTATG TACGGGTTAA AAATGGCGGC ATAGACGGTA AAGSTGCATT
	801	GTGTGTCAGC AGCAATACAC AAATCAATGT TTCAGGCAGC CTGAAAAACT
	851	CAGCGAGAT TCGAGGGCGC AATGCGCTTA TTATCAATCA CGATACGCTA
	901	GACATATCG GTGGGGGTAT TCATGCGTA AATACAGCG TTACGCGAC
60	951	ACAGACATCT AATAATATTG GCGGCACTT CTTCGCGGAA CAGACATAT
	1001	TGCTCATATG GGGTAAACAC ATCAACACCC AAAGACGCGT CAGACGAGT
	1051	CRAAATGCAC AAGGTAGCAG CACCTACCTA GACCAGATGG CAGGTATTTA

1101	TATCACAGGC	AAAGAAAAG	GTGTTTATG	AGCGCAGGCA	GGCAAGACA
1151	TCAACATCAT	TGCCGGTCAA	ATCAGCAATC	AATCAGATCA	AGGGCAAAAC
1201	CGGCTGCAGG	CAGGACGCGA	CATTAACTGT	GATACGGTAC	AAACCGGCAA
1251	ATATCAAGAA	ATCCAATTG	ATGCCGATAA	CCATACCATC	CGAGGTTCAA
1301	CGAACGAAGT	CGGCAGCAGC	ATTCAACAAA	AAGGCGATGT	TACCCtatTG
1351	TACGGGAATA	TCTCAATGTC	CAAAAGCTGCC	GAAGTCGGCA	GGCGAAAAGG
1401	CACACTTGCC	GTGTATGCTA	AAAATGACAT	TACTATCAGC	TACGGCATCC
1451	ATCGCGCGCA	AGTGTATGAT	GTCGCAAAAC	ATACAGCGAG	AAGCGCGCGC
1501	GTATTAATAT	TAGTCATTAC	CGATAAAGCC	CAAGCTCATC	ACGAACATGC
1551	TCAAGCAGC	ACCTTTGAAG	CAAGCAAGT	TGTATTGCGA	CGAGGAAACG
1601	ATGCCAACAT	CCTTGGCAGT	AATGTATTAT	CGGATAATGG	CACCCGGATT
1651	CAAGCAGGCA	ATCATGTTGC	CATTGTGTACA	ACCCAATCT	AAAGCCAAAG
1701	CGAAACCTAT	CATCAAAACC	AAAATCAGG	ATTGATGAGT	GCAGGTATCG
1751	GCTTCACTAT	TGGCAGCAGG	ACAAACACAC	AAGAAAACCA	ATCCCAAAGC
1801	AACGAACATA	CAGGCAGTAC	CCTAGGCAGC	CTGAAAGGCG	ATACCAACAT
1851	TGTTGCAAGC	AAACACTACG	AACAAACCGG	CAGCAACGTT	TCCAGCCCTG
1901	AGGGCAACAA	CCTTATCAGC	ACGCAAAAGT	TGGATTGTGG	CGCAGCACAA
1951	AACCAATTAA	ACAGCAAAAC	CACCCAACCC	TACGAAACAA	AAGGCTTAAC
2001	GGTGGCATTG	AGTTGCGCCG	TTACCGATT	GGCAACAAAC	CGGATTGCGG
2051	TAGCACACAA	AGCAGCAAAC	AAGTCGACAA	AAGCAAAAC	GACCGGTTTA
2101	ATCGCATGGC	GGCTGCAAT	CGAGTTGGC	AGGCGTATCA	AACAGGCAAA
2151	GGCGCAAAA	ACTTAG			

This corresponds to the amino acid sequence -SEQ ID 522; ORF117ng-1>:

25	1	LLVQTEKDGL	HNEQTFGEKK	VFSENGKLN	YWRARRKGD	ETGHIREQNYT
	51	LPEEITRDIS	LGSFAYESH	KALSRHAFSG	GTELPGNSRD	NIRTAKSNGI
	101	SLPYTFNSFT	PLPGSSLYII	NFANKGYLVE	TDFRFANYRQ	WLGS DYMLGS
	151	LKLDPNLHK	RLGDGYEQR	LINEQIAELT	GHRRLDGYQN	DEEQFKALMD
	201	NGATAARSMN	LSVGIALSAE	QAQAQLTSDIV	WLQKEVKLP	DGGTQTVLMP
	251	QVYVRVKNKG	IDKGALLSG	SNQINVSQS	LKNSGTLAGR	NALINTDTL
30	301	DNIGGRHAQ	KSAVTATQDI	NNIGGILSAE	Q'LLNAGNIN	INNQSATKSS
	351	QNAQGSSTYL	DRMAGIYITG	KEKGVLAQAQ	GKDNIIAQI	ISNQSDQOGT
	401	RLQAGRDINL	DTVQTKGYQE	IHPDNRHTI	RGSTNEVSGS	IGTKGVDVTL
	451	SGNNINAKRA	EYSGAKGTLA	VYKNDITIS	SGHAGVVD	ASGHTGRSGG
	501	GNKLYTIDKA	QSHETAGQS	FTEGKQVQLQ	AGNDANLGS	NWISDNGTIR
35	551	QAGNHVIRIGT	TQTQSQSEY	HQTKQSLGMS	AGIGTITGSK	TWTQENQSGS
	601	NEHTGSTVGS	LKGDTTIVAS	KHYEQTGSNV	SPFEGNNLIS	TQSDMIGAAQ
	651	NQLNSKTTQT	YEQKGLTVAF	SPSFVTLAQQ	AIATAHKAAN	KSDKAKT'AL
	701	MPNRLPMQVG	RPIKQAKAH	T*		

ORF117ng-1 shows the same 90% identity over a 230aa overlap with ORF117. In addition, it

40 shows homology with a secreted *N. meningitidis* protein in the database:

	gi 2623258 (AF030941)	putative secreted protein [Neisseria meningitidis]Length = 2273
	Score =	604 bits (1541), Expect = e-172
	Identities =	325/678 (47%), Positives = 449/678 (65%), Gaps = 22/678 (3%)
45	Query: 1	LLVQTEKDGLHNEQTFGEKKVFSENGKLNLYWRARRKGDHDETGHIREQNYTLPEEITRDIS 60
		L+V T + L N++T G K + ++ G LH Y R +KG D TG+ Y E++ I
	Sbjct: 739	LIVGTPEESALDNDETLGTNTI-TDKGDLIRHYRHKKGRDSTGYRSRSPYEPAPFVS-SIR 796
50	Query: 61	LGSFAYESHKALSRHAPSGTGLPGNSRDNIRTAKSNGISLPYTFNSFTPLPGSSLYII 120
		+G AY+ + AF Q +++P + + NGI +T L P SSL+ I
	Sbjct: 797	MGTSYAKGY-----APQQAQSDIPGTV---VFWAENGHPTI-----LPNSSFLAI 840
55	Query: 121	NPANKGYLVETDPRFANYRQWLGS DYMLGSLKLDPNLHKRLGDGYEQRLINEQIAELT 180
		P NKGYL+ETDP F +YR+WLGS YML +L+ DPN++HKRLGDGYEYQ+L+HNEQIA+LT
	Sbjct: 841	APNNKGYLIEIDPAFTDYRWKLGSGYMLAALQDPDNIHKRLGDGYEYQKLINEQIAKLT 900
	Query: 181	GHRRLDGYQNDEEQFKALMDNGATAARSMNLSVGIALSAEQQAQLTSDIVWLQKEVKLP 240
60		G+RRLDGY NDEEQFKALMDNG T A+ +L+ GIALSAEQ A+LTSDIVWL + V LP
	Sbjct: 901	GYYRLDGYTNDEEQFKALMDNGITIAKELQLTPGIALSAEQVARLTSDIVWLENETVTL 960
	Query: 241	DGGTQTVLMPQVYVRVKNKGIDKGALLSGNSTQINVSQSLKN-SGTIAGRNALINTDT 299
		DG TQTVL P+VYVR + ++GALLSGS I SG++N G IAGR ALI+H
	Sbjct: 961	DGTTQTVLKPVKVYVRPKDMNGQCALLSSVVDIG-SGAENRGGLIAGREALINLQN 1019
65	Query: 300	LDNIGGRHAQKSAVTATQDINNIGGILSAEQ'LLNAGNINNQSATKSSQNAQGSSTY 359

-307-

- 5
10
15
20
25
- Sbjct: 1020 + N+ G + + A DI N G I AE LLL A NNI ++S +S+QN QGS
1KNLGGDLGGKNIFAAAGSDITNTGSI-GAENALLLKASNNIESRSETRNQNQGSVRN 1078
- Query: 360 LDRMAGIYITGKEKGVLAQAQAGKDINIIAGIOSNISNODOGOTRLQAGRINDLDTVQTKYQ 419
+ R+AGIY+TG++ G + AG +I + A +++NOS+ GOT L AG DI DT + Q
- Sbjct: 1079 IGRVAGIYLTGRQNGSVLLDAGNNIVLTASELTNQSEDDGOTVLNAGGDIRESDTTIGISRNQ 1138
- Query: 420 EIHFDADNHTIRGTSNEVGSSTOTKGDVTLTSGNNLNKAAAEVGSAGKTLAVYAKNDITI 479
FD+DN+ IR NEVGS+I+T+G+I+L + + +AAEVGS +G L + A DI +
- Sbjct: 1139 NTIFPDSNVYIRKEQNEVGSSTIRTRGNLSNAKGDIRIRAAEVGSEQGRKLAAAGDIRK 1198
- Query: 480 SSGIHAGOVDDASKHTGRSGGGKLVITDKAQSHHETAQSTFEGKOVVLQAGNDANILG 539
+G + +DA K+TGRSGGG K +T + + + A S T +GK+I+L +G D + G
- Sbjct: 1199 EAGKAHTETEDALKYTRSGGGGIKQKMTIRHLKNNQNGQAVSGTLDGKEIILVSGRDITVTG 1258
- Query: 540 SNVISDNGTRIQAGNHVRIGTTTOTOSSSETYHOTOKSGLM-SAGIGFTIGSKTNTQENQS 598
SN+I+DN T + A N++ + +T+S+S ++ +KSGLM S GIGFT GSK +TQ N+S
- Sbjct: 1259 SNIIADNHTIILSAKNNIVLKAETRSRGAEMNKKEKSGLMGSGGIGFTAGSKKDTQTNRS 1318
- Query: 599 QSNHTGTVSGSLKGDTTIVASKHYEOTGSNVSSPEGNNLISTOSMDIAGAAQNQLNSKTT 658
+IT S VGLS G+T I A KHV QTGS +SSP+G+ IE+ + I RAQN+ + ++
- Sbjct: 1319 ETVSHTESVVGSLNGNTLISAGKHHTYQTGSTISSPQGVGSISSGKISIDAAQNRYSQESK 1378
- Query: 659 QTYEOKGLTVAFSSPVD 676
Q YEOKG+TVA S EV +
- Sbjct: 1379 QVYEOKGVTVAISVFPVN 1396

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 63

- 30 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 523>:

1 ATGATTTACA TCGTACTGTT TCTAGCTGTC GTCTCGCGCG TTGTGCGCTA
51 CAACATGTAT CAGGAAAACC AATACCGCAA AAAATCGGCG GACCACTTCG
101 GACACTCCGA CAAGATGCC CTGCTCAACA GCAAAACGAG CCATGTCGCG
151 GACGCAAAAC CGTCCGCGCG GTCACTCATG ATGCCGAAC CCACACCGCG
35 201 GGTCAAAAAA ACGGCAAAAC CCAAGACCC CGYCATCGCG AACCTGCAAG
251 AACAGGATCG CGTCTACATC GCCAAGCAGA AACAGCGCAA AGCCTCCCGC
301 TTCAAAACCG AAATCGAAAC CGCCTTGGAA GAAAGCGGCA TTATCGGCAA
351 CTCGCGCCAC ACGCTTTCCG AACCCCAAA CCGACATTCC GCAACGAAAC
401 CTGCGGACGC GTGCGCAAAA CTTGCAACCG TTCCGCAAAC ACCTGCAAAA
40 451 CCGCTGATTA CGTCAAGA ACTGTCAAAA GTCGAATTAT CTGCGTTTGA
501 CGTGGCGATC GACTTCATCT CCTAT...

This corresponds to the amino acid sequence <SEQ ID 524; ORF119>:

1 MIYIVLFLAV VLVAVYNNY QENQYRKVR DOFGHSDKDA LLNSXTSHVR
45 51 EGKSGGGSVM MEKQFAVKK TAKFDQFMR NLQEQDAVY AKQKQAKASP
FKTEIEALE ESOIIGNSAH TVSEFQTHS ATKFDASAK PAPVPQTPAK
151 FLITLKELSK VELSWDFVRI DFISY...

Further work revealed the complete nucleotide sequence <SEQ ID 525>:

1 ATGATTTACA TCGTACTGTT TCTAGCTGTC GTCTCGCGCG TTGTGCGCTA
51 CAACATGTAT CAGGAAAACC AATACCGCAA AAAATCGGCG GACCACTTCG
50 101 GACACTCCGA CAAGATGCC CTGCTCAACA GCAAAACGAG CCATGTCGCG
151 GACGCAAAAC CGTCCGCGCG GTCACTCATG ATGCCGAAC CCACACCGCG
201 GGTCAAAAAA ACGGCAAAAC CCAAGACCC CGCATCGCG AACCTGCAAG
251 AACAGGATCG CGTCTACATC GCCAAGCAGA AACAGCGCAA AGCCTCCCGC
301 TTCAAAACCG AAATCGAAAC CGCCTTGGAA GAAAGCGGCA TTATCGGCAA
351 CTCGCGCCAC ACGCTTTCCG AACCCCAAA CCGACATTCC GCAACGAAAC
55 401 CTGCGGACGC GCGCGCAAAA CTTGCAACCG TTCCGCAAAC ACCTGCAAAA
451 CCGCTGATTA CGTCAAGA ACTGTCAAAA TCGGAATTAT CTGCGTTTGA
501 CTGCGGACGC GACTTCATCT CCTATCGC GCTGACGAA GCGCAAGAC
551 TGCACGCACT CGCGCGCCTT TCCAACCGCT GCGCTACACA GATTGTCGCG
60 601 TGCACCATGG ACGACCATTT CCAGATTGCC GAACCCATCC CGGCGCATCCG

-308-

5 651 CTATCAGGCA TTTATCSTGG GTAITCAGGC AGTCAGCCGC AACGGACTTG
701 CCTCGCAGGA AGAACTCTCC GCATTCAACC GCCAGGTGGA CGCATTCGCA
751 CAAAGCATGG GCGGTCAAGC GCTGCACACC GACCTTGCCG CCTTTATCGA
801 AGTGGCTTCC GCACTGGACG CATTTCTGCG GCGGCTCGAC CAGACCATCG
851 CCATCCATTT GGTTTCCCGC ACCAGCATCA GCGGCGTAGA ACTGCGTTCC
901 GCCGTAACGG GCGTGGGTTT CGTTTGTGAA GACGACGGCG CGTTCCACTA
951 TACCGACACG TCGGGCTCGA CCATGTTCTC CATCTGCTCG CTCACAACG
1001 AGCGGTTTAC CAACGCCCTT TTGACCAACC AGTCCTACAA AGGCTTCAGT
1051 ATCGGCTCGA ACATCCGAGA CTCTCGGCGA GCGCAAAA CCTTCGACGA
1101 TTTGTTTATG GATTTGGGCG TACGCTGTCT CGGCGAGTTG AACCTGAATC
1151 TGGTCAACGA CAAATAGGAA GAAGTTTCGA CCCAATGGCT CAAAGACGTG
1201 CGCACTTATG TATTGGCGCG TCAGTCCGAG ATGCTCAAAG TCGGTATCGA
1251 ACCGGGCGCG AAAACGSCAT TGCCTCTGTT CTCTTAA

This corresponds to the amino acid sequence <SEQ ID 526; ORF119-1>:

15 1 MIYIVLFLAV VLAUVAYNMY QENQYRKKVR DQFGHSDKDA LLNSKTSHVR
51 DGKPSGGSVN MPKPQPAVKK TAKPQDPAMR NLQEQDAVYI AKQKQAKASP
101 FKEIEITALE ESGIIGNSAH TVSEPQTGHS APKPADAPAK PAPVPQTPAK
151 PLITLKELSK VELPWFQVRF DFISYIALTE AKELHALPRL SNRCRYQIVG
201 CTMDHFQIA EPIGIRYQA FIVGIQAVSR NGLASQEELS AFNRQVDAFA
251 QSMGGTLLHT DLAAFLIVAS ALDAFCARVD QTIAIHLVSP TSISSVELRS
301 AVTGVGFVLE DGAHFHTYT SGTSMFICS LNNPEFTNAL LDNSYKGFES
351 MLLDIPHSPA GEKTFDLEFM DLAVLSGLQ NLNLVNDKME EVSTQWLKDV
401 RTYVLARQSE MLKVGIEFQG KTAALRFS*

Computer analysis of this amino acid sequence gave the following results:

25 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF119 shows 93.7% identity over a 175aa overlap with an ORF (ORF119a) from strain A of *N. meningitidis*:

		10	20	30	40	50	60
30	orf119.pep	MIYIVLFLAVLAVLAVVAYNMYQENQYRKKVRDQFGHSDKDALLSXTSHVRDQKPSGGSVN					
	orf119a	MIYIVLFLAAVLAVLAVVAYNMYQENQYRKKVRDQFGHSDKDALLSKTSHVVDGKPSGGPTM	10	20	30	40	50
		70	80	90	100	110	120
35	orf119.pep	MPKPQPAVKKTAKPQDPXMRNLQEQDAVYIAKQKQAKASPFKEIEITALEESGIIGNSAH					
	orf119a	MPKPQPAVKKTAKSDPAMRNLQEQDAVYIAKQKQAKASPFKEIEITALEESGIIGNSAH	70	80	90	100	110
		130	140	150	160	170	
40	orf119.pep	TVSEPQTGHSATKPADASAKPAPVPQTPAKPLITLKELSKVELSWFQVDFDFISY					
	orf119a	TVPEPQTGHSAPKADAPAKVPVPQTPAKPLITLKELSKVELPWFQVDFDFISYIALTE	130	140	150	160	170
45		190	200	210	220	230	240
	orf119a	AKELHALPRLSNRCRYQIVGCTMDHFQIAEPIGIRYQAFIVGIQAVSRNGLASQEELS					

The complete length ORF119a nucleotide sequence <SEQ ID 527> is:

50 1 ATGATTTACA TCGTACTGTT CTOCGCCGCC GTCCCTGCCG TTGTGCGCTA
51 CAATATGTAT CAGGAAAACC AATACCGCAA AAAAGTGGCG GACCAGTTTCG
101 GGCACTCCGA CAAAGATGCC CTGCTCAACA GCRAAACACG CCATGTCCCG
151 GACGGCAAAAC CGTCCGCGCG GCCAGTCATG ATGCCGAAAC CCAACCCGCG
201 GGTCAAAAAA ACGGCAAAAT CCAAGACACC GCCATGGCG AACCTGCAAG
55 251 ACGAGGATGC CTGTCAATC GCCAAGCAGA AACAGCGAAA AGCCTCCCGC
301 TTCAAAACCG AAATCGAACC CGCCTTGGAA GAAAGCGGCA TTATGGGCAA
351 CTCGCCACAC AAGCTTCAAG ACCCGAAC CTGACATTCC CACCAAAAC
401 CTCGCCACCG GCGGCAAAAC CCGTCTCCCG TCCGCGAAC GCGGCAAAA
451 CCGCTGATTA CGCTCAAAGA GCTGTGAGG GTGAGCTGC CCGTGGTTGA
501 CGTGCSCCTC GACTTCATCT CTTATATCGC GCTGACCGAA CCAAAAGAAC
60 551 TGCAACGCACT GCGCGGCCTT TCCAACCGCT GCGGCTACCA GATTGTGCGC
601 TGCAACCATG ACGACCAATT CCAGATTGCC GAACCATCC CCGGCAATCG

5	651	CTATCAGGCA	TTTATCGTCC	GTATTGAGGC	AGTCAGCCGC	AACGGACTTG
	701	CCTCGCAGGA	AGAACTCTCC	GCATTCACCC	GCCAGGTGGA	TGCATTGCGA
	751	CACAGCATCG	GGGTCAGAC	GCTGCACACC	GACCTTGCCG	CCTTATTGCA
	801	AGTCGCTTCC	GCATCGGACG	CATTCTGGCG	GCGCGTCGAC	CAGACTATCG
	851	CCATCGATTT	GCTTTCCCGC	ACCAGCATCA	GCGCGGTAGA	ACTCGGTTCC
	901	GCGGTAAACG	GCGTGGGTTT	CGTTTGGGAA	GACGACGGCG	CCTTCCACTA
	951	TACCGACACG	TGCGGCTCGA	CCATGTTCTC	CATCTGCTCG	CTCACACACG
	1001	AGCGTTTATC	CAATCCCTTT	TTGCGACACC	AGCTCTATTA	AGCGTTCATC
	1051	ATGCTGCTCG	ACATCCCGCA	CTCTCCGGCA	GCGCAAAAAC	CCTTCGACGA
10	1101	TTTGTTTATG	GATTTCCGGG	TAQCCCTGTC	GCGCGAGTTC	AACTCAATC
	1151	TGCTCAACGA	CAAAATCGAA	GAAGTTTCGA	CCCAATGGCT	CAAAAGACCTG
	1201	CGCACTTATG	TATTGGCTCG	TCAGTCCGAG	ATGCTCAAA	TGGTATCGA
	1251	ACCGGGGGCG	AAAACCGCAT	TGCGCTGTGT	CTCCTAA	

This encodes a protein having amino acid sequence <SEQ ID 528>:

15	1	MIYIVFLAA	VLAVVAYNMY	QENQYRKKVR	DQFGHSDKDA	LLNSKTSVHR
	51	DKFSGGCPVM	MPKQPAVKK	TAQSQDPAMR	NLQEQDAVYI	AKQKQAKAS
	101	FKTEIETALE	ESGIIGNSAH	TVPEPQTGHS	APKPADAPAK	PVPVPQTFAK
	151	PLTLTKELSK	VELFVDFVRF	DFISYIALTE	AKELHALPRL	SNRCRYQIVG
	201	CTMDHFEQIA	EPFPIGRYQA	FIVGIQAVSR	NGLASQEELS	AFNRQVDAFA
20	251	HSMSGQTLHT	DLAAFIEVAS	ALDAFCARVD	QTIAHLVSP	PTISISGVELRS
	301	AVTGVGFEVLE	DDGAFHYTDT	SGSTMFSICS	LNNEPFTNAL	LDNQSYKGS
	351	MLDIPHSFA	CEKTFDDLFM	DLAVRLSGQL	NLNLVNDKME	EVSTQWLKDV
	401	RTYVLARQSE	MLKVGIEPGG	KTALRLFS		

ORF119a and ORF119-1 show 98.6% identity in 428 aa overlap:

25		10	20	30	40	50	60
	orf119a.pep	MIYIVFLAAVLAVVAYNMYQENQYRKKVRDQFGHSDKDALNSKTSVHRDQKPSGGSPVM					
	orf119-1	MIYIVFLAAVLAVVAYNMYQENQYRKKVRDQFGHSDKDALNSKTSVHRDQKPSGGSPVM					
		10	20	30	40	50	60
30		70	80	90	100	110	120
	orf119a.pep	MPKQPAVKKRTAKSQDPAMRNLQEQDAVYIAKQKQAKASPFKTEIETALEESGIIGNSAH					
	orf119-1	MPKQPAVKKRTAKSQDPAMRNLQEQDAVYIAKQKQAKASPFKTEIETALEESGIIGNSAH					
35		70	80	90	100	110	120
	orf119a.pep	TVPEPQTGHSAPKPADAPAKPVPVPTPAKELITLKLKSKVELPWFDFVRFDFISYIALTE					
	orf119-1	TVPEPQTGHSAPKPADAPAKPVPVPTPAKELITLKLKSKVELPWFDFVRFDFISYIALTE					
40		130	140	150	160	170	180
	orf119a.pep	AKELHALPRLSNRCRYQIVGCTMDHFEQIAEPFPIGRYQAFIVGIIQAVSRNGLASQEELS					
	orf119-1	AKELHALPRLSNRCRYQIVGCTMDHFEQIAEPFPIGRYQAFIVGIIQAVSRNGLASQEELS					
45		190	200	210	220	230	240
	orf119a.pep	AFNRQVDAFAHSMGGQTLHTDLAFAIEVASALDAFCARVDQTIAHLVSPPTISISGVELRS					
	orf119-1	AFNRQVDAFAHSMGGQTLHTDLAFAIEVASALDAFCARVDQTIAHLVSPPTISISGVELRS					
50		250	260	270	280	290	300
	orf119a.pep	AVTGVGFEVLEDDGAFHYTDTSGSTMFSICS LNNEPFTNALDNQSYKGSMLLDIPHSFA					
	orf119-1	AVTGVGFEVLEDDGAFHYTDTSGSTMFSICS LNNEPFTNALDNQSYKGSMLLDIPHSFA					
55		310	320	330	340	350	360
	orf119a.pep	GEKTFDDLFMDLAVRLSGQLNLNLVNDKMEVSTQWLKDVSTYVLARQSEMLKVGIEPGG					
	orf119-1	GEKTFDDLFMDLAVRLSGQLNLNLVNDKMEVSTQWLKDVSTYVLARQSEMLKVGIEPGG					
60		370	380	390	400	410	420
	orf119a.pep	GEKTFDDLFMDLAVRLSGQLNLNLVNDKMEVSTQWLKDVSTYVLARQSEMLKVGIEPGG					
	orf119-1	GEKTFDDLFMDLAVRLSGQLNLNLVNDKMEVSTQWLKDVSTYVLARQSEMLKVGIEPGG					
65		370	380	390	400	410	420

```

orf119a.pep      KTALRLFSX
                  |||||
orf119-1         KTALRLFSX

```

5 Homology with a predicted ORF from *N.gonorrhoeae*

ORF119 shows 93.1% identity over a 175aa overlap with a predicted ORF (ORF119ng) from *N. gonorrhoeae*:

10	orf119.pep	MIYIVLFLAAVLAVVANNYMQEQYKKVKVRDQGHSDKDALINSKTSHVRRGKPSGGVSM	60
	orf119.ng	MIYIVLFLAAVLAVVANNYMQEQYKKVKVRDQGHSDKDALINSKTSHVRRGKPSGGVSM	60
15	orf119.pep	MPKQPQAVKTKAKPQPMXRNLEQDAVYIAKQKQAKASPFKTEIETALEESGIIGNSAH	120
	orf119.ng	MPKQPQAVKPKAKPQDSMRNLEQDAVYIAKQKQAKASPFKTEIETALEEIGIIGNSAH	120
	orf119.pep	TVSEPTQGHSAKPADAKAPVPPQTPAKPLILTKELSKVLEWDFVDFDISY	170
	orf119.ng	TVSEPTQGHSAKPADAPAKVPPVPTQTPAKPLILTKELSKVLEWDFVDFDISYALTE	180

The complete length ORF119ng nucleotide sequence <SEO ID 529> is:

20	1	AAGATTTCAG	TGCTACTGTT	CTCTGGCGCC	GTCTCTGGCG	TTGTGGCGTCA
	51	CAATATGPTAT	CAGGAAAACC	AATAGTCGAA	AAAGTGGCTG	GACAGCTTCG
	101	GACCTCCGGA	CAAGATGTCC	GTCTCAACA	CGMAAACACG	CCATGTCGCG
	151	CGCGGCAAAC	CGTCCGGGCG	CGCAGTCTAT	ATGCGGAAAC	CCCAACCGCG
	201	GGTCAAAAAA	CGGGCAACCA	CCCAAGACTC	CGCAGTGGCG	AACCTCGAAG
25	251	AACAGGATGC	CGTCTACATC	AGCGACAGA	AGCCGCAAAA	AGCCTCCCGC
	301	TTCAAAACGC	AAATCGAAAC	CGCCTTGAAA	GAAATCGCCA	TTATCGGCAA
	351	CTCGCGCCAC	ACCGTTTCGG	AACCCGAAC	CGGACATTCG	GCAACGAAC
	401	CTCGCGACGC	CGCGGCAAAA	CCCGTTCCCG	TTCGCGAAAC	CGCGGCAAAA
	451	CGCTGTGATTA	CGCTCAAGAA	GTGNTGCGAG	GTAGGAGCTCG	CTCTGGTTTGA
30	501	CGTGGCTCTC	gACTCTCATCT	CCCTATTCGC	GTAGCCCGGA	GCCTAAGAAC
	551	TCGACGACCT	CGCGGCGGCT	TCGACGCGCT	TCGACGCGCT	GA TTTGCGCT
	601	TGCGACATTC	AGGACATTCG	CCAGATTCGC	GACCCCTACT	CGGGATTCGC
	651	CTATCAGCGA	TTTATCGTGG	GTATCTCAGC	AGTCACGCGC	AACGGACTTG
35	701	CTCTCGCAGG	AGAACTCTCC	CGCAGCGGGA	CGCAGGCGGA	GCATTTCGCA
	751	CAAGACATGG	CGGGTCAGAC	CGTCCACGCG	GACCTTGGCG	CTTTATTGCA
	801	AGTGCGTTCG	CGACTGGAAG	CATTCTCAGC	CGGGCTGCAG	CAGACCATCG
	851	CGATCCATTT	GGTTTGGCGG	ACCGAGTATCA	CGGGGCTAGA	ACTGGGTTCG
	901	CGCTCAACGG	CGCGGGGTTT	CGTCTTGAAA	GACGACGCTG	GGTTCACATA
40	951	TACCGACAGG	TGCGGGCTCG	CGATGTTCTC	CATCTGCTCG	CTCAACGAAG
	1001	AGCGCTTTAC	CAATCGCCTT	TTGGGACAAC	AGGCTCTACA	AGGCTTCAGT
	1051	ATGTCGCTCG	ACATCCCGCA	CTTCGACACG	GGGCGAAAAA	CTCTGACGGA
	1101	TTTGGTTCGG	TTTGGTTCGG	TATCGCTATC	ATCTGCTGCT	ATCTGCTGCT
	1151	TGCTCAAGCA	CAAAATGGAA	GAAGTTTCGA	CCCAATGGCT	CAAGACGACT
	1201	CGCATTATG	TATTTGGGCG	TGCGCTCCAG	ATGTCCTAAG	GTGTTATGCA
45	1251	ACCGGGGCGC	AAAACGGCCG	TCGCGCTTCT	TTTCATBA	

This encodes a protein having amino acid sequence <SEQ ID 530>:

	1	MIYIVFLVLA	VLAVVANNMY	QENQYRKRV	DQFGHSDKDA	LINSKTSVHR
	51	DQPSGGPVM	MKFPQAVKK	PAKPDQSMR	NIQQDQAVYI	AKQQAQASV
50	101	FKTEIETALE	EIGIIGNSAH	TVSEFPQGH	AKFPADPAK	PVPFPQPAK
	151	PLTIKELSK	VELPFDVRF	DFISYLAITE	HLALHLPAL	SNRCRQIVG
	201	CTMDHFGTIA	EPIFIRYQA	FVIGIQAVSR	NGLASAEELS	AFNRNDGALS
	251	QSMGGQQLA	DLAETFEVS	ALDAFCQVR	QTAJHLVSP	TSIGVDEFA
	301	AVFGVQVLA	SGCAETFTT	SGCTMFLSC	LNHNSVQVH	LNHNSVQVH
	351	MLLDIPHSA	GKTKFTDLM	DLAVRLSGQL	LNHNLVNKME	EVSQTQWLKV
55	401	RTYVLARQE	MLKVGIEFGG	KTALRFLS*		

ORF119ng and ORF119-1 show 98.4% identity over 428 aa overlap:

60

	10	20	30	40	50	60
orf119ng	MIYIVLFLAAVLAVVAYNQYQNRKKYVDQFGHSDKDALLNSKTSHVDRGKPSGGSPVM					
orf119-1	MIYIVLFLAAVLAVVAYNQYQNRKKYVDQFGHSDKDALLNSKTSHVDRGKPSGGSPVM					
	10	20	30	40	50	60

-311-

		70	80	90	100	110	120
5	orf119ng	MPKFPQPAVKKPAKPDQSAMRNLEQDDAVYIAKQKAKASPFKTEETALEEIGIISNAH					
	orf119-1	MPKFPQPAVKKTAQPDQPMARNLEQDDAVYIAKQKAKASPFKTEETALEEIGIISNAH					
		70	80	90	100	110	120
10	orf119ng	130	140	150	160	170	180
	orf119-1	TVSEPTGHSAPKPADAPAKPVPVPTPAKPLITLKEISKVELPWFDFVDFTSYIALTE					
		130	140	150	160	170	180
15	orf119ng	190	200	210	220	230	240
	orf119-1	AKELHALPRLSNRCRYIVGCTMDHFIABPIPIGIRYQAFIVGIQAVSRNGLAQEELS					
		190	200	210	220	230	240
20	orf119ng	250	260	270	280	290	300
	orf119-1	AFNRQADAFQSMGGQTLHTDLAAFIIVASALDAFCARVDQTTIAHLVSPISISGVELRS					
		250	260	270	280	290	300
25	orf119ng	310	320	330	340	350	360
	orf119-1	AVTGVGFVLEDDGAFHYTDTSGMTSFICSINNEPFTNALLDNQSYKGSMLLDIPHSFA					
		310	320	330	340	350	360
30	orf119ng	370	380	390	400	410	420
	orf119-1	GEKTFDDLFDLAVRLSGQLNLVNDKMEEVSTQWLKDVRTYVLARQSEMLKVGIEPGG					
		370	380	390	400	410	420
35	orf119ng	429					
	orf119-1	KTALRLFSX					
		KTALRLFSX					
40	orf119ng						
	orf119-1						

Based on this analysis, including the presence of a putative leader sequence in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

45 Example 64

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 531>

	1	.GC	CGCGGCAGC	GCACGGGAAGA	TTTCTTCATG	AACACACGCG	ACAC	ATCAG
	51	CG	AGTAGTATC	GAAGGCACCA	CCGGTAGCAT	GAGCTCGTGT	ATTTCCTCCA	
	101	TC	GCCTCGTAT	TTTCTTTGGTA	GTCGCGCGCA	TCGCGGTGAT	GAACATCATG	
50	151	CT	GGTGTCCG	TTACCGGAGCG	CACCAAGAAG	ATCGGCATAC	GGATGGCAAT	
	201	CG	CGCGCGCGG	CGCGGCAATA	TTTtGCAGCA	GTTTTTGATT	GAGCGGGTGT	
	251	TA	ATCTGCGT	CATCGGGCGT	TTGGTCGGCG	TGGGTTTGTC	CGCGCGCGCT	
	301	AG	CGCTCGTG	TCAATCATTT	TGTAACCGAC	TTCCCGATGG	ACATTTCCGC	
	351	CA	TGTCGCGT	ATCGCGCGCG	TCGCCTGTT	GACCGGAATC	GGCATCGCGT	
55	401	TC	GGCTTTAT	GCCTGCCAAT	AAAGCAGCCA	AACTCAATCC	GATAGACGCA	
	451	TT	GGCACAGG	ATTGA				

This corresponds to the amino acid sequence <SEQ ID 532; ORF134>:

	1	.AR	HGTEDFFM	NNSDXIRQIV	ESTTGTMKLL	ISSIALISLV	VGGIGVMNIM	
	51	LV	SVTERTEKE	IGIRMAIGAR	RGNIQQQLI	EAVLICVIGG	LIVGVLISAAV	
60	101	SL	VFNHFVTD	FFMDISAMSV	IGAVACSTGI	GIATGFMPAN	KAALKLNPIDA	
	151	LA	QD*					

Further work revealed the complete nucleotide sequence <SEQ ID 533>:

	1	ATGTCGGTGCG	AGACGATATT	GGCGACACAA	ATGGGTTGCG	TTCTCGAAGT
	5	51	GCTCGGCATCT	ATCATCGGTA	TGGCTCGGTG	GTCGATCGAT
	101	GCATGCTGTTT	GCACATATTTA	ATGCTTGAAG	ACATCATGAT	GATGAGGACG
	151	AACACCATCTA	GCAGTATTTCC	GCGGCGCGCG	TTCCGGGACA	GCGCAGCGGG
	201	CAGGATTAAAC	ACCTCTGACCA	TAGACGACGC	AAAAATCATC	GCCAAACAAA
	251	GCTAGTCTTGG	TTCCGCGCAG	CGATCGATCT	CGAGCGGCGG	CACGCTGACT
	301	TACGCGACGA	CGCGCTGAC	CGCTCGGTTT	TACGCGTGG	GGAACACATA
	351	TTTGACCACTA	CGCGACTGTA	CGCGGAATCT	GGGGCGGCTG	TTTGACGAAA
10		401	AGCATGTGAA	AGAGAACGCG	GCGCTCGTGG	TCGATGACCA
	451	GCAAACTCTT	TTGCGGCTCT	GGATCGCTTG	TGACGAGCTA	TTTGGTCAAA
	501	GAAGCGCTGA	CGGCGGCTGA	CGGCGGCTGA	GAAAGCGGCTG	GAAAGCGGCTG
	551	TGCGCAATTC	CGAGCTGCTG	ATGCTTTGGT	CGCCGCTATG	GACGGTGATG
	601	CACCAATCTA	CAGCGGAGCA	CCACACCAAC	TCCTCATCAC	TCAAAATGAT
15		651	AGACATGTCG	ATGCTCCRAG	TTCGCGAAAT	AGGGCTGACC
	701	AGGCGCGGCA	CGGACCGGAA	GATTTCTTCA	CGACAGATCT	CGACAGATCT
	751	AGCGAGATAG	TCGAAAGAGC	CAGCTGGCTG	ATGAGGCTCT	TGATTTCTCT
	801	ATCGGCGCTG	ATTTCATTGG	TACGCGGGCG	CATCGGCGTG	ATGAACATCA
	851	TCGCTGCTGT	CTATCCAGCG	CGACCAACAG	AAATCGGCTC	ACGGATGGCA
20		901	ATCGGCGGTC	GCGCGGCGAA	GATTTTGGAG	CAGTTTITTA
	951	GTATATCTCG	GTCATCGGCG	GTTTGGTGGG	CTCGGGTGTG	TTCGCGCGCG
	1001	TCAGGCTGCT	TTTCAATCAT	TTTGGATGAG	ACTCTCCGAT	GGACATTTCC
	1051	GCGATCTGCT	GGTGGCTGCT	GGTGGCTGCT	GGTGGCTGCT	TGGGCTGCTG
	1101	GTTGGCGCTT	ATGCTGCGCA	ATAAAGCAGC	CBAATCTCAN	CCGATAGAGC
25		1151	CATGGGCACTA	GGATTTGTA		

30

1	MSYQVAILHGR	MRSLILPMIGI	IIGTASVVSQV	VALGNQSGQK	ILEDISSIGIGI
51	NTSL:IFKRG	FDRRASGRIK	PIIITDDAKII	AVKQ	AMQSSGGLITL
101	YKFDLSDSL	YGVGEQFQDV	RLGIKLTGRL	LDAGQASAT	GVVIVDQKVA
151	DKFLDADSL	PKTGLKFFDP	LTGVVGMKKD	ENAFNDSGIV	MLMSPTVTV
201	HQITGESHTN	STVTKIKDNA	TSIVQAEKGLT	DLKKARQETE	DFEMFNNDST
251	QRVESITGVI	MKLLISSIAL	NILQVAVVGE	MMIVLSVSTE	RTKEIGIRMA
301	ITGARGNIIQ	QFLTEAVLTIC	VGVGIVLGVV	SAVLSVFNHN	FTVTFDPMDIS
351	RMVSVGAVC	QSTGLGAEFC	MPNKAZKAIN	PIITALAOC	

35 Computer analysis of this amino acid sequence gave the following results:

Homology with the hypothetical protein o648 of *E. coli* (accession number AE000189)

ORF134 and o648 protein show 45% aa identity in 153aa overlap:

40	Crf134: 2	RHGTEFDPMNNSDXIRQIVETGT+TGMKXXXXXXXXXXXXVVGIGVGNMIMLVSTERTKE	61
	o648: 496	RHG+DFF N D + + V E TT + + VVGIGVGNMIMLVSTERTKE	
45	Crf134: 62	GRMAIGARRGNIXQFLIEAXXXXXXXXXXXXXXXXXXFNHFDTDFMDISAMSV	121
	o648: 556	GIRMA+GAR + + QFLIEA F+ + + S + + +	
		GRIMAVGARSSVLOQFLIEAVLVLCLVGALGDTLSLLIAFTLQLFLPGWEIGFSPALL	615
	Crf134: 122	GAVACSTGIGIAGFGFMANKAKALINPDAALQ 154	
	A CST G FG+ + PA AA+L+P DAA+ +		
o648: 616	LAFICSTVIGLFGFMARNALDPVDAL+R		

50 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF134 shows 98.7% identity over a 154aa overlap with an ORF (ORF134a) from strain A of *N. meningitidis*:

55

orf134.pep orf134a	<div style="margin-bottom: 10px;"> 10 20 30 ARHGTEDFFMNSDXIRQIVESTTGTMKLL</div> <div>GESHTNITVTKIKDNANTOVAEKGLTDLLKARHGEDFFMNSDSIRQIVESTTGTMKLL</div> <div style="margin-top: 10px;"> 210 220 230 240 250 260</div> <div style="margin-top: 20px;"> 40 50 60 70 80 90</div>
---------------------------	--

orf134.pep	ISSIALISLVGGIGVMNIMLVSVTERTKSIGIRMAIGARRGNIXQQFLIEAVLICVIGG
orf134a	ISSIALISLVGGIGVMNIMLVSVTERTKSIGIRMAIGARRGNILQQFLIEAVLICVIGG
	270 280 290 300 310 320
orf134.pep	LVGVGLSAAVSLVFNHFVTDFFMDISAMSVIGAVACSTGIGIAFGMPANKAAKLNIPDA
orf134a	LVGVGLSAAVSLVFNHFVTDFFMDISAMSVIGAVACSTGIGIAFGMPANKAAKLNIPDA
	330 340 350 360 370 380
orf134.pep	LAQDX
orf134a	LAQDX

The complete length ORF134a nucleotide sequence <SEQ ID 535> is:

1	ATGTCGGTGC AAGCAGTATT GCGCACAAAT ATGCGTTTCG TTCTGACGAT
51	GCTCGGCATC ATCATCGGTA TCGCTTCGGT TGCTCCGCTC GTCCGATTGG
101	GCAACGGTTC CGAGAAAAAA ATCCTTGAAG ACATCAGTTC GATAGGGACG
151	AACACCATCA GCATCTTCCC AGGGCGCGCG TTGCGGACGA GCGCGACGGG
201	CAGATTAA ACCTGACCA TCAGACACGC AAAAATCATC GCCAAACAAA
251	GTCTAGTTGC TTCCGCGACG CCATGACTT CGAGCGCGCG CACGCTGACT
301	TACCGCAATA CGACCTGAC CGCTTCTTTG TACGCTGGG GCGAACATA
351	TTTCGACGTG CGCGGGCTGA AGCTGGAAC GGGCGCGCTG TTTGACGAAA
401	ACGATGTGAA AGAAGACGCG CAGGTCGTGC TCATCGACCA AATGTCAAA
451	GACAACTCTT TTGCGGACTC GGATCGGTGT GGTAAACCA TTTTGTTCAG
501	GAAACGCCCC TTGACGCTCA TCGGCGGTGAT GAAAAAGAC GAAAACGCTT
551	TCGGCAATTC CGACGTGCTG ATGCTTTTGT CGCCTATAC GACGCTGATG
601	CACCAATATCA CAGGCGAGAG CCACACCAAC TCCATCACCG TCAAAATCAA
651	AGACAATGCC AATACCCAGG TTGCGGAAA AGGGCTGACC GATCTGCTCA
701	AAGCGCGGCA CGGCACGGAA GATTTCCTTA TGAACACAG CGACAGCATC
751	AGGCASATAG TCGAAAGCAC CACCGCTACG ATGAAGCTCG TGATTTCTCT
801	CAFCGCCCTG ATTTCATTGG TAGTCGCGCG CATCGCGGTG ATGAACATCA
851	TCGCTGCTTC CATTACCGAG CGCACCAAG AATCGGCAT ACGAATGCCA
901	ATCGCGCGCG CGCGCGGCAT TATTTTGCAG CATTTTTCGA TTGAGCGCGT
951	GTTAATCTCG GTCATCGCGG GTTTGTCGG CGTGGGTTG TCGCGCGCGG
1001	TCAGCCTCGT GTTCAATCAT TTTGTAACCG ACTTCCGAT GGACATTTC
1051	GCCATGTCGG TCATCGCGCG GGTGCGCTGT TCGACCGGAA TCGGCATCGC
1101	GTTGCGCTTT ATGCTGCGCA ATAAAGCAGC CAAACTCAAT CGATAGATG
1151	CATTGGCGCA GGATTGA

This encodes a protein having amino acid sequence <SEQ ID 536>:

1	MSVQVLAHK MRSLLTMLGI IIGIASVVSV VALNGSQKK ILEDISSIGT
51	NTISIFPGRG FGDRRSGRIK TLTIDDAKII AKQSYVASAT PMTSSGGTLT
101	YRNTDLTASL YGVGEQYFDV RGLKLETGRL FDENDVKEDA QVVVDQNVK
151	DKLFADSDPL GKTILFRKRP LTVIVGMKRD ENAFNGSDVL MLWSPYTTVM
201	HQITGESHTN STVVKIKDNR NTVQAEKGLT DLLKARHGT EDEFFMNSDSI
251	RQIVSESTCT MLLISSIAL ISLVGVGSL MNIMLVSVTE RKEIGIRMA
301	IGARRGNILQ QFLIEAVLIC VIGGLVGVGL SAASLVFNHF FVTDFFMDIS
351	AMSVIGAVAC STGIGIAFGF MPANKAAKLN PIDALAQA

ORF134a and ORF134-1 show 100.0% identity in 388 aa overlap:

orf134a.pep	MSVQVLAHKMRSLLTMLGIIIGIASVVSVVALNGSQKKILEDISIGTNTISIFPGRG
orf134-1	MSVQVLAHKMRSLLTMLGIIIGIASVVSVVALNGSQKKILEDISIGTNTISIFPGRG
orf134a.pep	FGDRRSGRIKTLTIDDAKIIAKQSYVASATPMTSSGGTLTYRNTDLTASLYGVGEQYFDV
orf134-1	FGDRRSGRIKTLTIDDAKIIAKQSYVASATPMTSSGGTLTYRNTDLTASLYGVGEQYFDV
orf134a.pep	RGLKLETGRLFDENDVKEDAQVVVDQNVKDKLFADSDPLGKTLILFRKRP LTVIVGMKRD
orf134-1	RGLKLETGRLFDENDVKEDAQVVVDQNVKDKLFADSDPLGKTLILFRKRP LTVIVGMKRD
orf134a.pep	ENAFNGSDVLMWSPYTTVMHQITGESHTNSITVVKIKDNANTQVAEKGLTDLLKARHGT
orf134-1	ENAFNGSDVLMWSPYTTVMHQITGESHTNSITVVKIKDNANTQVAEKGLTDLLKARHGT

	orf134a.pep	DDFMNNSDSIRQIVESTTGTMKLLISSIALISLVVGGIGVMNIMLVSVTERTKEIGIRMA	
5	orf134-1	DDFMNNSDSIRQIVESTTGTMKLLISSIALISLVVGGIGVMNIMLVSVTERTKEIGIRMA	
	orf134a.pep	IGARRGNILQQFLIEAVLICVIGGLVGVGLSAAVSLVFNHFVTFDFPMDISAMSVIGAVAC	
	orf134-1	IGARRGNILQQFLIEAVLICVIGGLVGVGLSAAVSLVFNHFVTFDFPMDISAMSVIGAVAC	
10	orf134a.pep	STGIGIAFGFMPANKAAKLNPIDALAQDX	
	orf134-1	STGIGIAFGFMPANKAAKLNPIDALAQDX	
<u>Homology with a predicted ORF from <i>N.gonorrhoeae</i></u>			
15	ORF134 shows 96.8% identity over a 154aa overlap with a predicted ORF (ORF134.ng) from <i>N.gonorrhoeae</i> :		
	orf134.pep	ARHGTEDDFMNNSDXIRQIVESTTGTMKLL	30
20	orf134ng	GESHTNSITVKIKDNANTRVAEKGAELELLKARHGTEDDFMNNSDSIRQIVESTTGTMKLL	264
	orf134.pep	ISSIALISLVVGGIGVMNIMLVSVTERTKEIGIRMAIGARRGNIXQQFLIEAVLICVIGG	90
	orf134ng	ISSIALISLVVGGIGVMNIMLVSVTERTKEIGIRMAIGARRGNILQQFLIEAVLICVIGG	324
25	orf134.pep	LVGVGLSAAVSLVFNHFVTFDFPMDISAMSVIGAVACSTGIGIAFGFMPANKAAKLNPIDA	150
	orf134ng	LVGVGLSAAVSLVFNHFVTFDFPMDISAAVIGAVACSTGIGIAFGFMPANKAAKLNPIDA	384
30	orf134.pep	LAQD 154	
	orf134ng	LAQD 388	

The complete length ORF134ng nucleotide sequence <SEQ ID 537> is:

	1	ATGTCGGTGC	AAGCAGTATT	GGCGCACAAA	ATGCGTTGCG	TTCTGACCAT
35	51	GCTCGGCATC	ATCATCGGTA	TGCGTTGCGT	TGCTTCGGTC	GTCGCGCTGG
	101	GCAACGGTTC	GCAGAAAAAA	ATCTCGGAAG	ACATCAGTTC	GATGGGGACG
	151	AACACCATCA	GCATCTTCCC	CGGGCGGGCG	TTGCGGACAC	GCGCGACGGG
	201	CAAAATCAAA	ACCCGTACCA	TAGACGACGC	AAAAATCATC	GCCAAACAAA
	251	GCTACGTTGC	CTCGGCCAAG	CCCATGACTT	CGAGCGGGCG	CACGCTGACC
40	301	TACCGCAATA	CGGACCTGAC	CGCTTCTTTG	TACGGTGTGG	GCGAACRAAT
	351	TTTGCACGTG	CGGGCGGTGA	AGCTGGAAC	GGGGCGGCTG	TTTGATGAGA
	401	ACGATGTGAA	AGAGAGACGC	CAAGTCGTGC	TGCTCGACCA	AAATGTCAAA
	451	GAACAACCTC	TTGCGGACTC	GGATCGGTGG	GSTAAACCA	TTTTGTTCAG
	501	GAACACGCC	TTGACCGTCA	TGCGCGTGAT	GAAGAAAGAC	GAAGACCTT
45	551	TGCGCAATTC	CGACGTGCTG	ATGCTTTGGT	CGCCCTATAC	GACGGTGATG
	601	CACCAAAATCA	CAGGCGGAGG	CCACACCAAC	TCCATCACCG	TCAAAATCAA
	651	AGACAATGCC	AATACCGGGG	TTGCGCAAAA	AGGGTGTGCC	GAGGTGCTCA
	701	AAGCACGGCA	CGGCACGGAA	GACTTCTTTA	TGAACAACAG	CGACAGCATC
	751	AGGCAGATGG	TGCAAAAGCAC	CACCGGTACG	ATGAAGCTGC	TGATTTCCCT
50	801	CATCGCCCTG	ATTTCATTGG	TAGTCGGGGG	CATCGGTGTG	ATGAACATTA
	851	TGCTGGTGTC	CGTTACCGAG	CGCACCAAG	AAATCGGGAT	ACGGATGGCA
	901	ATCGCGCGCG	GGCGGCGCAA	TATTTTGCAG	CAGTTTTTGA	TTGAGGCGGT
	951	GTTAATCTGC	ATCATCGGAG	GCTTGTGGGG	CGTAGGTTTG	TCCGCGCGCG
	1001	TCAGCGCTCG	GTTCAATCAT	TTTGTAAACG	ATTTCCCGAT	GGACATTTTG
	1051	GCGGCATCCG	TTATCGGGCG	GCTCGGCTGT	TGACCGGGAA	TCCGATCGCG
55	1101	GTTCCGCTTT	ATGCGTGGCA	ATAAGGCAGC	CAAACTCAAT	CCGATAGATG
	1151	CATTGGCGCA	GGATTGA			

This encodes a protein having amino acid sequence <SEQ ID 538>:

	1	MSVQAVLAHK	MRSLLTMLGI	IIGIASVVSV	VALNGSQKK	ILEDISSMGT
60	51	NTISIFPGRG	FGDRRSQKIK	TLTIDDAKII	AKQSYVASAT	PMTSSGGTLT
	101	YRNTDLTASL	YGVGEQYFDV	RLGKLETGR	FDENDIDQNVK	QVVVIDQNVK
	151	DKLFADSDPL	GKTLILFRKP	LTIVGVMMKD	ENAFGNSDVL	MLWSPYTTVM
	201	HQITGESHTN	SITVKIKDNA	NTRVAEKG	ELLKARHGTG	DDFMNNSDSI
	251	RQMVESTTGT	MKLLISSIAL	ISLVVGGIGV	MNIMLVSVTE	RTKEIGIRMA
	301	IGARRGNILQ	QFLIEAVLIC	IIGGLVGVGL	SAAVSLVFNH	FVTFDFPMDIS

351 AASVIGAVAC STGIGIAFGF MPANKAAKLN PIDALAQD*

ORF134ng and ORF134-1 show 97.9% identity in 388 aa overlap:

```

5   orf134ng      MSVQAVLAHKMRSLLTMLGIIIGIASVSVVALNGNSQKKILEDISMGTNTISIFPGRG
      orf134-1    MSVQAVLAHKMRSLLTMLGIIIGIASVSVVALNGNSQKKILEDISMGTNTISIFPGRG

10  orf134ng      FGDRRSGKIKTLTIDDAKIIAQSYVASATFMTSSGGTLTYRNTDLTSLYGVGEQYFDV
      orf134-1    FGDRRSGKIKTLTIDDAKIIAQSYVASATFMTSSGGTLTYRNTDLTSLYGVGEQYFDV

15  orf134ng      RGLKLETGRFLFENDVKEDAQVVVIDQNVKDKLFADSDPLGKTLFRKRPLTVIGVMKKD
      orf134-1    RGLKLETGRFLFENDVKEDAQVVVIDQNVKDKLFADSDPLGKTLFRKRPLTVIGVMKKD

20  orf134ng      ENAFGNSDVLMLWSPYTTVMHQITGESHTNSITVKIKDNANTRVAEGLAELLKARHGT
      orf134-1    ENAFGNSDVLMLWSPYTTVMHQITGESHTNSITVKIKDNANTQVAEGLTDLKARHGT

25  orf134ng      DFFMNNSDSIRQMVSTTGTMKLLISSIALISLVVGGIGVMNIMLVSVTERTKEIGIRMA
      orf134-1    DFFMNNSDSIRQIVESTTGTMKLLISSIALISLVVGGIGVMNIMLVSVTERTKEIGIRMA

      orf134ng      IGARRGNILQQFLIEAVLICIGGLVGVGLSAASVSLVFNHFVTFPMDISAAVIGAVAC
      orf134-1    IGARRGNILQQFLIEAVLICVIGGLVGVGLSAASVSLVFNHFVTFPMDISAAVIGAVAC

      orf134ng      STGIGIAFGFMPANKAAKLNPIDALAQDX
      orf134-1    STGIGIAFGFMPANKAAKLNPIDALAQDX

```

30 ORF134ng also shows homology to an *E. coli* ABC transporter:

```

sp|P75831|YBJZ_ECOLI_HYPOTHETICAL_ABC_TRANSPORTER_ATP-BINDING_PROTEIN_YBJZ_gi1
(AE000189) o648; similar to YBBA_HAEIN_SW: P45247 [Escherichia coli] Length =
648
35   Score = 297 bits (753), Expect = 6e-80
      Identities = 162/389 (41%), Positives = 230/389 (58%), Gaps = 1/389 (0%)

Query: 1   MSVQAVLAHKMRSLLTMLXXXXXXXXXXXXXNGNSQKKILEDISMGTNTISIFPGRG 60
      M+ +A+ A+KMR+LLTML          +G+ +++ +L DI S+GTNTI ++PG+
40  Sbjct: 260 MAWRALAANKMRTL TMLGIIIGIASVSVVVGDAKQMVLAIRISGTNTIDVYPGKD 319

Query: 61   FGDRRSGKIKTLTIDDAKIIAQSYVASATFMTSSGGTLTYRNTDLTSLYGVGEQYFDV 120
      FGD  + L DD I KQ +VASATP S L Y N D+ AS GV YF+V
Sbjct: 320 FGDDDPQYQALKYDYLIAIQKQPVASATFAVSGNLRRLRYNNVDVAASANGVSGDYFNV 379

45  Query: 121 RGLKLETGRFLFENDVKEDAQVVVIDQNVKDKLFAD-SDPLGKTLFRKRPLTVIGVMKK 179
      G+ G F++ + AQVVV+D N + +LF +D +G+ IL P VIGV ++
Sbjct: 380 YGMTFTSEGNTFNQQLNGRAQVVVLDNSTRKQLFPHKADVVEGIVLGNMFPARVIGVAEE 439

50  Query: 180 DGNAFGNSDVLMLWSPYTTVMHQITGESHTNSITVKIKDNANTRVAEGLAELLKARHGT 239
      ++ FG+S VL +W PY+T+ ++ G+S NSITV++K+ ++ AE+ L LL RHG
Sbjct: 440 KQSMFGSSKVLRLVWLPYSTMSGRVMQSWLNSITVRVKEGDSAAEQQLTRLLSLRHGK 499

Query: 240 EDFFMNNSDSIRQMVSTTGTMKXXXXXXXXXXXXXVGGIGVMNIMLVSVTERTKEIGIRM 299
      +DFF N D + + VE TT T++          VVGGIGVMNIMLVSVTERT+EI GIRM
55  Sbjct: 500 KDFFTWNMDGV LKTVKTTTTLQLFELTLVAVISLVVGGIGVMNIMLVSVTERTREI GIRM 559

Query: 300 AIGARRGNILQQFLIEXXXXXXXXXXXXXXXXXXXXXFNHFVTFPMDISAAVIGAVA 359
      A+GAR ++LQQFLIE          F+ + + S +++ A
60  Sbjct: 560 AVGARASDVLQQFLIEAVLVCLVGGALGTLTSLIAFTLQLFLPGWIEIGFSLALLAFL 619

Query: 360 CSTGIGIAFGFMPANKAAKLNPIDALAQD 388
      CST GI FG++PA AA+L+P+DALA+
Sbjct: 620 CSTVTGILFGWLPAARNAARLDPVDALARE 648

```

Based on this analysis, including the presence of the leader peptide and transmembrane regions in the gonococcal protein, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 65

- 5 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 539>:

```

1      ..GGGACGGGAG CGATGCTGCT GCTGTTTAC GCGGTAACGA T.CTGCCTTT
51     GGCCACTGGC GTTACCCTGA GTTACACCTC GTCGATTTTT TTGGCGGTAT
101    TTTCTCTCTC GATTTTGAAA GAACGGATTI CCGTTTACAC GCAGCGCGTG
151    CTGCTCCTCT GTTTTGGCCG GCTGCTATTG CTGCTTAATC CCTCGTTCGG
10     CAGCGGTCAG GAAACGGCGG CACTCGCCGG GCTGGCGGGC GGCGCGATGT
251    CCGGCTGGGC GTATTTGAAA GTGCGCGAAC TGCTTTTGGC GGGCGAACC
301    GGCTGGCGGC TCGTGTTTTA CTTTCGCTG ACAGGTGTGG CGATGCTGCT
351    GGTGTGGGCG AGCTGACCCG GCTGCGCACG CCGTCTCTTT CCATCGGCAG
15     TTATCTGCTC GTGCACTGGC GTGCGCGGGC TGAATGGACA ACTGTGATG
451    ACGCGCGGCT ACAAAGTCGG GCACAAATTC ACGGTGCGCT CGCTTCTCTA
501    TATGACCGCT GTTTTTCGG CTCTGTCTGC CGCATTTTTT CTGGGCGAAG
551    AGCTTTTCTG GCAGGAATA CTCGGTATGT GCATCATCAT CCGTACGGGT
601    ATTTTGA

```

This corresponds to the amino acid sequence <SEQ ID 540; ORF135>:

```

1      ..GTGAMLLLFY AVTILPLATG VTLSTSSIF LAVFSLILK ERISVYTOAV
51     LLLGFAGVVL LLNFSFRSQ ETALAGLAG GAMSQWAYLK VRELSLAGEP
101    GWRVVFYSLV TGVAMSSVMA TLTGWTLSF FSAVYLSGIV VLSALIAQLSM
151    TRAYKVGDKF TVASLSYMTV VFSALSAFF LGEELFWOEI LMCIIISAV
201    F*

```

- 25 Further work revealed the complete nucleotide sequence <SEQ ID 541>:

```

1      ATGGATACCG CAAAAAAGA CATTTTAGGA TCGGCTGGA TGCTGCTGGC
51     GCGCGCTCGC TTTACATTA TGAACCTATT GATTAAAGAG GCATCGGCAG
101    AATTTCOCCT CGGACGCGCT GAATTGGCTT TTTGCGCAT GCTCTTTTCA
151    ACGGTTGGCG TCGGGCTGCG CGCCGATATG CGTGGGACA mCTTCGCGAC
201    CGCCCATTTG AAAAAACCACT TAAACCGCAG TATGCTCGGA ACGGGGCGCA
251    TGCTGCTGCT GTTTTACGCG GTAACCGCATC TGCTCTTGGC CACTGCGGTT
301    ACCCTGAGTT ACACCTCGTC GATTTTTTTG GCGGTATTTT CCTTCTGAT
351    TTTGAAAGAA CGGATTTCGG TTTACACGCA GCGGCTGCTG CTCTTGGTT
401    TTGCGCGCGT GGTATTGCTG CTTAATCCCT CGTTCGCGAG CGGTACAGAA
15     ACGCGCGCAC TCGCGGGCT GCGGGCGGCG GCGATGTCGG GCTGGGCGTA
501    TTTGAAAGTG CGCGAAGTGT CTTTGGCGGG CGAACCGCGC TGGCGGCTCG
551    TGTTTTACCT TTCCGTGACA GGTGTGGCGA TGCTCTCGGT TTGGGCGAAG
601    CTGACCGGCT GGCACACCTT GTCTTTTCCA TGCGCAGTTT ATCTGTGCTG
651    CATGCGGCTG TCGGCTGTA TTGCCCCAAT CTGCGATGAG CGCGCTTACA
40     AGCTCGGCGA CAATTCACG GTTGGCTCGG TTTCCTATAT GACCGCTGTT
751    TTTTCCGCTC TGCTCGCGC ATTTTCTTGG GCGGAAGAGC TTTTCTGSCA
801    GGAATACTCT GGTATGTGCA TCATCATCCT CAGCGGTATT TTGAGCAGCA
851    TCGCCCCAC TGCCTTCAAA CAGCGGCTGC AATTCCTGTT CGGCCAAGA
901    TAA

```

- 45 This corresponds to the amino acid sequence <SEQ ID 542; ORF135-1>:

```

1      MDTAKKDILG SGWMLVAAAC FTIMNVLIKE ASAKFALSG ELVEWRMLFS
51     TVALGAAAVL RRDXTFRPHN KNHLNRMVG TGAMLLLFYA VTHLPLATGV
101    TLTSTSSIFL AVFSEFLIKE RISVYTAQVL LLGFAGVVLN LNPFSRSGQE
151    TAALAGLAGG AMSQWAYLKV RELSLAGEFG WRVVFYLSVT GVMSSVWAT
201    LTGWTLSLF SAVYLSGIV SALIAQLSMT RAYKVGDKET VASLSYMTV
50     FSAALSAFFL GEELFWOEIL GMCIILSGI LSSIRPTAFK QRLQSLFRQR
301    *

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF135 shows 99.0% identity over a 197aa overlap with an ORF (ORF135a) from strain A of *N.meningitidis*:

```

5      orf135.pep      10      20      30
                        GTGAMLLLFYAVTILPLATGVTLSTSSIF
orf135a      STVALGAAAVLRDRTFRTPHNKNHNSRMVGTGAMLLLFYAVTILPLATGVTLSTSSIF
                        50      60      70      80      90      100

10     orf135.pep      40      50      60      70      80      90
                        LAVFSLIILKERISVYTOAVLLLGFGAGVLLNPSFRSGQETAALAGLAGGAMSGWAYLK
orf135a      LAVFSLIILKERISVYTOAVLLLGFGAGVLLNPSFRSGQETAALAGLAGGAMSGWAYLK
                        110     120     130     140     150     160

15     orf135.pep      100     110     120     130     140     150
                        VRELSLAGEPGWRVVFYLSVTGVAMSSVWATLTGWHTLSFSPSAVYLSICIGVSALIAQLSM
orf135a      VRELSLAGEPGWRVVFYLSVTGVAMSSVWATLTGWHTLSFSPSAVYLSICIGVSALIAQLSM
                        170     180     190     200     210     220

20     orf135.pep      160     170     180     190     200
                        TRAYKVGDKFTVASLSYMTVVFSAASAFFLGEELFWQELGMCIIISAVFX
orf135a      TRAYKVGDKFTVASLSYMTVVFSAASAFFLAEELFWQELGMCIIISGIISSIRPTAF
                        230     240     250     260     270     280

orf135a      KQRLQSLFRQXX
                        290     300

```

30 The complete length ORF135a nucleotide sequence <SEQ ID 543> is:

```

1      ATGGATACCG CAAAAAAGA CATTTAGGA TCGGGCTGCA TGCTGGTGGC
51     GCGCGGCTCG TTTACCATTA TGMACGTATT GATTAAAGAG GCATCGGCAG
101    AATTGTGCCCT CGGCAGCGCG GAATTGGTCT TTTGGCCCAT GCTGTTTCCA
151    ACCGTTGGCG TCGGGGCTCG CGCCGTATTG CGTCGGGACA CTTCCGCAC
201    GCCCATATGG AAAAACCACT TAAACGCAGC TATGCTCGGG ACGGGGGCGA
251    TGCTGCTGCT GTTTTACGCG GTAACGCATC TGCCTTTGGC CACGGGGGTT
301    ACCCTGAGTT ACACCTCGTC GATTTTTTGG GCGGTATTTT CTTCTCTGAT
351    TTTGAAAGAA CGGATTTCCG TTTACACGCA GCGGCTGCTG CTCCTTGTTT
401    TTGCGGCGCT GGTATTGCTG CTTAATCCCT CGTTCGCCAG CGGTGAGGAA
451    ACGCGCGCAC TCGCGGGGCT GCGGGGCGCG CGGATGTCCG GTCGGGCGTA
501    TTTGAAAGTG CGCGAAGCTG CTTTGGCGGG CGAACCCGCG TGGCGCGCTG
551    TGTTTACACT TCCCGTGACA GGTGTGGGCA TGTCATCGGT TTGGCGAGCG
601    CTGACGCGCT GGCACACCTT GTCCTTTCCA TCGGCAGGTT ATCTGTCGGT
651    CATCGCGCTG TCGCGGCTGA TTGCCCATC GTGCATGACG CGCGGCTGCA
701    AAGTCGGCGA CAAATTCACG GTTGCCCTCG TTTCCTATAT CACCGCTGCT
751    TTTTCCGCTC TGTCGCGCCG ATTTTTCCTG CGCGAAGAGC TTTTCTGGCA
801    GGAATACTCT GGTATGTGCA TCATCATCCT CAGCGGTATT TTGAGCAGCA
851    TCGGCCCCAC TGCTTTCARA CAGCGGCTGC AATCCCTGTT CGGCAARAAG
901    TAA

```

50 This encodes a protein having amino acid sequence <SEQ ID 544>:

```

1      MDTAKKDILG SGWMLVAAAC FTIMNVLIKE ASAKFALGSG ELVFWRMFLS
51     TVALGAAAVL RDRDTRTPHW KNHNSRMVGT GAMLLLFYA VTHLPLATGV
101    TLSYTSISIFL AVFSLILKE RISVYTOAVL LLGFAGVLLN LNPFSRSGQE
151    TAALAGLAGG AMSGWAYLKV RELSLAGEPG WRVVFYLSVT GVAMSSVWAT
201    LTGWHTLSFP SAVYLSICIGV SALIAQLSMT RAYKVGDKFT VASLSYMTVV
251    FSALSAFFLL AEELFWQEL GMCIIILSGI LSSIRPTAFK QRLQSLFRQR
301    *

```

ORF135a and ORF135-1 show 99.3% identity in 300 aa overlap:

```

60     orf135a.pep      MDTAKKDILGSGWMLVAAACFTIMNVLIKEASAKFALGSGELVFWRMFLSTVALGAAAVL
                        |||
orf135-1      MDTAKKDILGSGWMLVAAACFTIMNVLIKEASAKFALGSGELVFWRMFLSTVALGAAAVL

```

	orf135a.pep	RRDTFRTPHWKNHLNRSVMGTGAMLLLFYAVTHLPLATGVTLSTYSISIFLAVVFSFLILKE
5	orf135-1	RRDXFRTPHWKNHLNRSVMGTGAMLLLFYAVTHLPLATGVTLSTYSISIFLAVVFSFLILKE
	orf135a.pep	RISVYTQAVLLGFAGVLLNPFSGQETAALAGLAGGAMSGWAYLKVRELSLAGEPG
	orf135-1	RISVYTQAVLLGFAGVLLNPFSGQETAALAGLAGGAMSGWAYLKVRELSLAGEPG
10	orf135a.pep	WRVVFYLSVTGVAMSSVWATLTGWHTLSFSPSAVYLSGIGVSALIAQLSMTRAYKVGDKFT
	orf135-1	WRVVFYLSVTGVAMSSVWATLTGWHTLSFSPSAVYLSGIGVSALIAQLSMTRAYKVGDKFT
	orf135a.pep	VASISYMTVVFVSALSAFFLAELFWQELGMCIIILSGILSSIRPTAFKQRLQSLFRQR
15	orf135-1	VASISYMTVVFVSALSAFFLGEELFWQELGMCIIILSGILSSIRPTAFKQRLQSLFRQR

Homology with a predicted ORF from *N.gonorrhoeae*

ORF135 shows 97% identity over a 201aa overlap with a predicted ORF (ORF135ng) from

20	<i>N.gonorrhoeae</i> :		
	orf135.pep	GTGAMLLLFYAVTXLPLATGVTLSTYSISIF	30
	orf135ng	STVTLGAAVLRDRFTFPHWKNHLNRSVMGTGAMLLLFYAVTHLPLATGVTLSTYSISIF	335
25	orf135.pep	LAVFSFLILKERISVYTQAVLLGFAGVLLNPFSGQETAALAGLAGGAMSGWAYLK	90
	orf135ng	LAVFSFLILKERISVYTQAVLLGFAGVLLNPFSGQETAALAGLAGGAMSGWAYLK	395
	orf135.pep	VRELSLAGEPGWRVVFYLSVTGVAMSSVWATLTGWHTLSFSPSAVYLSGIGVSALIAQLSM	150
30	orf135ng	VRELSLAGEPGWRVVFYLSATGVAMSSVWATLTGWHTLSFSPSAVYLSGIGVSALIAQLSM	455
	orf135.pep	TRAYKVGDKFTVASISYMTVVFVSALSAFFLGEELFWQELGMCIIISAVF	201
35	orf135ng	TRAYKVGDKFTVASISYMTVVFVSALSAFFLGEELFWQELGMCIIISAAF	506

An ORF135ng nucleotide sequence <SEQ ID 545> was predicted to encode a protein having amino acid sequence <SEQ ID 546>:

	1	MPSEKAFRRH	LRTASFOGLH	LHHFHQKVKG	CGIIGFGIHI	FPTLLPAAQG
	51	ILDIQLGLFR	IDFAALAVYR	RTQVDFIHTV	IDGIASDAQF	SEVVQILRRL
40	101	NLGHFTDTHL	IAQARRFIAD	FGNIRPMRRG	EAKTEFCRCR	FDGIDGIGHD
	151	FRQCCHINRL	APGDKCRNGK	RDKVFFHTRH	YNQVCLEKTN	CSARKIKFRH
	201	QKQAKTHSTS	LAARFTIRPS	LSQRPFMDTA	KDKILDSGWM	LVAAACFTVM
	251	NVLIKERASAK	FALGSGELVF	WRMLFSTVTL	GAARVLRARD	FRTPHWKNHL
45	301	NRSVMGTGAM	LLLFYAVTHL	PLTTGVTLST	TSSIFLAVFS	FLILKERISV
	351	YTCQVLLLEF	AGVLLLNFS	FSGQEPAL	ASLAGGAMSG	WAYLKVRELS
	401	LAGEPGWRVY	FYLSNTGVAM	SVWATLTGW	HTLSFSPSAVY	LSGIGVSALI
	451	AQLSMTRAYV	VGDKFTVASL	SYMTVVFVSAL	SAFFLGEEL	FWQELGMCII
	501	IIISAAF*				

Further work revealed the following gonococcal sequence <SEQ ID 547>:

50	1	ATGGATACCG	CAAAAAAGA	CATTTTAGGA	TCCGGCTGGA	TGCTGGTGGC
	51	GGCGGCTCG	TTCAACGCTTA	TGAACGTATT	GATTAAGAG	GCATCGGCCA
	101	AATTTGCCCT	CGGCAGCGGC	GAATTTGTCT	TTTGGCGCAT	GCTGTTTTTC
	151	ACCGTTACGC	TCCGTGCTGC	CGCCGATTAT	CGCGCGCACA	CCTTCGCCAC
	201	GCCCCATTGG	AAAAACCACT	TAAACCGCAG	TATGCTCGGG	ACGGGGGCGA
55	251	TGCTGCTGCT	GTTTACCGCG	GTAACGACAT	TGCGTTTTCG	ACACCGCGGT
	301	ACCGTGAGTT	ACACCTCGTC	GATTTTTCTG	CGCGTATTTT	CCTTCCTGAT
	351	TTTGAAGAAG	CGGATTTCCG	TTTACACACA	CGCGGTGCTG	CTCTCTTGTT
	401	TTGCGCGCGT	GCTATTCCTG	CTTATCCCTC	CGTTCGCGAG	CGCTACGACA
	451	CCGCGCGCAT	TCCCGCGGCT	CGCGCGCGCG	CGGATGTCG	GCTGGCGCTA
	501	TTTGAAGAGT	CGCGAAGCTG	CTTTGGCGGG	CGAACCCGCG	TGGCGCGCTC
60	551	TGTTTTTACT	TTCCGCACCC	GGCCTGGCGA	TGTCGTGggt	tgggggagag
	601	Ctgaacggct	ggCAGAcct	GTCCTTcca	tgggagctgt	ATCgtCGGG

5

651	CATCGCGCGTG	tcgcgcgCtgA	TTGCGCAATC	GtcgatgAcy	cGCGccttaa
701	AAATCGCGGGA	TAATTCACGG	TTGTCCCTCGC	tttctctATat	cttctctCGTC
751	TTTTTCGCGCC	TGCTCTCGCGC	ATTTTTTCtg	ggcgcgaagac	tttTtctgCGA
801	GGGATTAATCTC	GGATATGTCGA	CTATTATctc	CAGCAGGCA	TTTGGAGCA
851	TCCGCCCCAT	TGCTTCTCAA	CAGCGCTGC	AAGCCCTCTT	CCGCCAAGA
901	Taa				

This corresponds to the amino acid sequence <SEO ID 548: ORF135ng-1>:

10

1	MDTAKKDILG	SGWMLVAAR	FTVMNVLIKE	ASAKFALSGG	ELVFWRLMFS
51	TVLTGYAAST	RVDFPTFKHW	KNHILNRSMVG	TGAMLLGVLLA	VTHLPLTTGGV
101	TLTSSAISFL	ARSTFLLIKE	IRISVTVQGL	LIGAGFVLLV	LNPSFSRGQGE
151	PALALAGLGG	AMSGWAYVKV	RELSLAQMG	NRVVFLYSAT	GVAMSSVMVAT
201	LTGWHITLSP	SAPVLSGGIV	SALIAQLMST	RAYVKGDKFT	VASLSYMTVTV
251	FSALSAAPFL	GEEFLFWQEL	GMCIILSLGI	LSSTRPFAFK	QRQLALFRQR
301	*				

15 ORF135ng-1 and ORF135-1 show 97.0% identity in 300 aa overlap:

20
25
30

orf135ng-1.pep	MDTAKKDIILGSGWMLVAAACFTIMNVILKEASAKFALGSGSELVFWRMFLFSTVTLGAAAVL
orf135-1	MDTAKKDIILGSGWMLVAAACFTIMNVILKEASAKFALGSGSELVFWRMFLFSTVTLGAAAVL
orf135ng-1.pep	RRDFTFTPHWKHNLNRSVMVGGAAMILLYFYATHLPLTTGVTLSTYSSIFLAVSFSLILKE
orf135-1	RRDFTFTPHWKHNLNRSVMVGGAAMILLYFYATHLPLATGVTLSTYSSIFLAVSFSLILKE
orf135ng-1.pep	RISVYTCQAVILLGFGAGVLLLNPSFRGGQEPAAAGLAGGAMSGWAYLKVRELSLAGEPG
orf135-1	RISVYTCQAVILLGFGAGVLLLNPSFRGGQETAAAGLAGGAMSGWAYLKVRELSLAGEPG
orf135ng-1.pep	WRVVFYLSATGVAMSSVWATLTGWHTLSFPFAVYLSGGIGVSLIAQLSMTRYKVGDKFT
orf135-1	WRVVFYLSVTVGAMSSVWATLTGWHTLSFPFAVYLSGIGVSLIAQLSMTRYKVGDKFT
orf135ng-1.pep	VASLSYMTVVFSLASAAFLPEGLGEWQELIGMCIILSGILSSIRIPAFKORQALFQR
orf135-1	VASLSYMTVVFSLASAAFLPEGLGEWQELIGMCIITLSGILSSIRIPAFKORQSLFQR

35 Based on this analysis, including the presence of several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 66

The following DNA sequence was identified in *N.meningitidis* <SEO ID 549>:

40	1	ATGAAGCGCG	GTATAGGCGT	CTTGTGCTGT	TATTCGGATG
	51	TTTGGGACAA	TTTGTCCGCA	TATGCTTCAA	TACAGTTCGG
	101	TGCTCTTCCA	GATTTTCGGG	ATATCTTTCTT	TCTGTATACA
	151	CTGCGCCGCA	TGGCCGAAAT	CGATTCCCA	TGTTCCGTGTC
45	201	GTCCTCTCTC	CGCTCATCTG	CGCGGAGTGT	TGCTGTATGT
	251	TAGGGGATGC	GTTGTCACAC	GRACATCCAG	TCGCTGATGT
	301	AGCGCAAAGC	TTTGGCTCTT	TGTGACATT	CGSGGTTCAT
	351	TATCTGCAC	ACCGTAAATA	TAAAGACGGT	CAAAATAAAT
	401	CACATATGTT	TCGAAATTTG	CGCTCTTCG	CGGCTTCGGA
	451	TTTACCATCT	GCAAAATCCA	AGGCGGAAAT	AAATGCGGCG
50	501	AAAGCTCGGG	CCAAAATGTT	TGATATGTTT	TAGGGAAGTTC
	551	CGSTTTATGC	CGTTTATGCA	TGCTGTATCA	TAAATTAAGG
	601	CATCATATGC	GTCTCTCAAC	TGATCGAGTA	CTGTGATCGC
	651	GTTTCTTGCC	KTGGGATCC	GATTGCGGTA	TGAAATATTC
	701	GATATG			mmjwyATTGC

55 This corresponds to the amino acid sequence <SEO ID 550: ORF136>:

1 MKRRIAVFVL FPQIIRVLGQ LLPKIVNTVP AHRMLFQIFG MFFFFIHQQY
51 LPGAIDSP CGIVFGALLF RHLPAHCLYG KAAVGDAVAH EHPADVNR

-320-

```
101 NANAFALFDI GQFAXFIVQH TVNIKTVKIN IVDPHMFANF AVFAVLEKRD
151 FDHGKIQGGN NAAAFPPKLA PKIFECFTGA FVGTVYREVC LFYIINDGIA
201 HESAPORVRY LEAPYCGELP SASDSLKSS XXSE*
```

Further work revealed the complete nucleotide sequence <SEO ID 551>:

5	1	ATGATGAAAGCG	GCGCATATAGC	GGTCTTGGTC	CTGTGTCGCG	AGATAATTCG
	51	AGTTTGGGGA	CACATCTGGT	CGAAATCTGT	CACATAGACT	CCGGCACATC
	101	GAGTGCTCTT	CCAGATGTTCT	GGGATGTTCT	TTTTCCTCAT	ACACCAGCAA
	151	TATCTGCCCG	GATGTCGCGA	AATCGATTCCT	CCATGCGGGA	TCGTGTCGCG
10	201	TGCGCTCCTC	TCCTGCATCT	CCGCGCGCGA	TTCGCGTGTAT	GGTAAAGCCG
	251	CCTGAGGGGA	TGCGCTTGCA	CAGCAATCAT	CAGTCGTGTA	TGTCGTCAC
	301	CGGAAGCGCA	ACGCTCTTGC	CTTGTGTACG	ATGCGTCAGT	TGCGCGGGTT
	351	CATTGTTCA	CACACCTGTA	ATAATAAGAC	CGTCARAAAT	AATATGCTCG
	401	ATCCACATAT	TTCGCGAAAT	TTGCGCGTCT	TGCGCGTCTT	GGAAAAGAGG
	451	GACTTGTACC	ATGGCAAAAT	CCAGAGCGGA	AATATGTGGG	CGGGGTCGCC
15	501	AAAAAGCTAG	CGTGCAGAAA	TATTGTGAAT	TTTTCGCGG	CGCTCTGCG
	551	GCACGGTGAT	CGCTGCTGTA	CTGCTGCTGT	TCGAGGAGAT	TCGAGGAGAT
	601	CCCTCATATT	CTGCTCCTCA	TTTATGACG	TTTATGACG	CACCTTCTGT
	651	CGCTTCTCT	CTCTCGGCAT	CCGATCTCGA	TTTGAAGAT	TCCAAATATT
	701	CGGAATAG				

20 This corresponds to the amino acid sequence <SEQ ID 552; ORF136-1>:

25

1	MMKRRIAVFV	LFPQIRVLG	QLLPKIVNTV	PAHRMLPQIF	GMFFFFIHQQ
51	YLPGIATIDS	PCGIVFGALL	FRHLPAHCLY	GKAAPDFAV	HEHFVADVVN
101	RNANAFALDP	IGQFAGFVIG	HTVNTKTKVI	NIVDPHMFAN	FAVFVALEKR
151	DFDHGIQQGV	NNAAAPFPKKL	APKISECTGP	AFVGTVYREY	CLFYIINDGI
201	AHHSAQTVR	YLAPYCGFL	SPASDSLDKS	SKSYE*	

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF136 shows 71.7% identity over a 237aa overlap with an ORF (ORF136a) from strain A of *N. meningitidis*:

30	orf136.pep	10	20	30	40	50	59
		MKRRIAVFVLFPQIIRVLGQLPKIVNTVAHRMLFQIGMFFFIHQVLEGIAIDIS					
	orf136a	10	20	30	40	50	60
		MMKRRIAVFLLMQIKIRILGQLPKIVNTVAHRMLFQIGMFFFIHQVLEGIAIDIS					
35	orf136.pep	60	70	80	90	100	110
		PCGIVFGALLFRHLPAHCLCYGKAAVGDVAHEHVPADVVRNANAFALFDIGQFAGFIVQ					
	orf136a	60	70	80	90	100	110
		PCGIVFGTLLFRHXSTHCLCYGKAAVGNVAHEHVPADVVRNANAFALFDIGQFAGFIVQ					
40	orf136.pep	120	130	140	150	160	170
		HTVNIKTVKINIVDPHEMFANFAVFAVLEKRLTGDSHGKIGGNNAAAFPKLAPKIFECFQV					
	orf136a	120	130	140	150	160	170
		HAINKTVKINIVDPHEMFANFAVFAVLEKRLTMKRSKXXMMRRASQKSRQKYNLVLRA					
45	orf136.pep	180	190	200	210	220	230
		AFVGTVYRVFLCYFIINDGTAHH---SAPQRYVLPAPYCGFLPSASDSGLSKSSXSEX					
	orf136a	180	190	200	210	220	230
		R---SPARTGLSACSXXMTSEPIISAPQRYVLPAPYCGFLPSASDSGLSKSSXSEX					

The complete length ORF136a nucleotide sequence <SEQ ID 553> is:

55

1	ATGATGAAGC	GGCGTATAGC	CGTCTTGGTC	CTGCTCATGC	AGAAATTCGG
5	GATTTGGGA	CACCTGGTGC	CGAAATCGT	CACTACAGT	CGCGACATC
9	GACGTCTCTT	CCAGATNTTC	GGGATGTTCT	TTTTCTCAT	ACACAGGAA
13	TAGTGCCTG	GAGTGCGGCA	AATCGTCCAT	CCATGCTTGG	
17	TCGCTCTCTT	TTCGGTCTCA	NGTCCAGCA	TGCTCGTAT	GGTAAAGCG
21	CACGTGGGAA	TGCGGTGACA	CACGACATC	CACTCGTGA	TGTGCTCAAC

-321-

301 CGGAACGCCAA ACGCTTTGCG CTTGTTGAC ATTGGTCAGT TCGCCGGGTT
 351 CATTTGTTGAC CAGCGCATAA ATGTAAAGAC CGTCAAAAAT AATATCGTCG
 401 ATCCACATAT GTTCGCAAAAT TTGCGCCTCT TCGCGCTCTT GGAAAAAAGG
 451 GCTTTGACCA TGGCAAAAAT TAAGGNGNNA NNGATCGGGC GCGTTCCTCA
 501 AAAAAGCTCG CGCAAAAAT ATTGGAATGT TTGCGGGCG CTTGCGCGGG
 551 CACGGTTTAC CGGTTTGTCT GCCTGTTCTA CATATAAAT GACGGAATCG
 601 CCGATCATAT CTGCTCTCA ACGTGTACGG TATCTGTTG CACCTTACTG
 651 CGGCTTTCTG CCGTGGCAT CGGATTGGGA TTTGAAAAAT TCCAAATATT
 701 CGGAATAG

10 This encodes a protein having amino acid sequence <SEQ ID 554>:

1 MMKRRIAVFV LLMQKIRILG QLLPKIVNTV PAHRMLFQXF GMFFFFIHQQ
 51 YLPGIAREIDS PCGIVFGTLL FRHXSTHCLY GKAAGNVA HEHPVADV
 101 RNANAFALFD IGQFAGFIVQ HAINVTKVKI NIVDPHMFAN FAXFAVLEKR
 151 ALTMASKXXX XMRRRSQKSS RQKYLNLVLA RSPARFTGLS ACST**MTES
 15 201 PIISAPQVRV YLFAPYCGFL PSASDSDLKS SKYSE*

ORF136a and ORF136-1 show 73.1% identity in 238 aa overlap:

		10	20	30	40	50	60
20	orf136a.pep	MMKRRIAVFVLLMQKIRILGQLLPKIVNTVPAHRMLFQXFGMFFFFIHQQYLPGLIAEIDS					
	orf136-1	MMKRRIAVFVLFQIIRVLGQLLPKIVNTVPAHRMLFQIFGMFFFFIHQQYLPGLIAEIDS					
		10	20	30	40	50	60
25	orf136a.pep	PCGIVFGTLLFRHXSTHCLYGKAAGNVAHEHPVADVNRNANAFALFDIGQFAGFIVQ					
	orf136-1	PCGIVFGALLFRHLFAHCLYGKAAGVDAVAHEHPVADVNRNANAFALFDIGQFAGFIVQ					
		70	80	90	100	110	120
30	orf136a.pep	HAINVTKVKINIVDPHMFANFAXFAVLEKRALTMASKXXXMRRSQKSSRQKYLNLVLA					
	orf136-1	HTVNKTVKINIVDPHMFANFVAVFAVLEKRDFDHGKIQGGNNAAFPKKLAPKIFECFTG					
		130	140	150	160	170	180
35	orf136a.pep	R---SPARFTGLSACSTXXMTESPIISAPQVRVYLFAPYCGFLPSASDSDLKSSKYSEX					
	orf136-1	AFVGTVYRVCLFYIINDGIAH---SAPQVRVYLFAPYCGFLPSASDSDLKSSKYSEX					
		190	200	210	220	230	
40							

Homology with a predicted ORF from *N.gonorrhoeae*

ORF136 shows 92.3% identity over a 234aa overlap with a predicted ORF (ORF136ng) from *N.gonorrhoeae*:

45	orf136.pep	MMKRRIAVFVLFQIIRVLGQLLPKIVNTVPAHRMLFQIFGMFFFFIHQQYLPGLIAEIDS	59
	orf136ng	MMKRRIAVFVLLMQKIRILGQLLPKIVNTVPAHRMLFQIFGMFFFFIHQQYLPGLIAEIDS	60
	orf136.pep	PCGIVFGALLFRHLFAHCLYGKAAGVDAVAHEHPVADVNRNANAFALFDIGQFAGFIVQ	119
50	orf136ng	PCGIVFGTLLFRHLFAHCLYGKAAGVDAVAHEHPVADVNRNANAFALFDIGQFAGFIVQ	120
	orf136.pep	HTVNKTVKINIVDPHMFANFVAVFAVLEKRDFDHGKIQGGNNAAFPKKLAPKIFECFTG	179
55	orf136ng	HTVNKTVKINIVDPHMFANFVAVFAVLEKRDFDHGKIQGGNNAAFPKKLAPKIFECFTG	180
	orf136.pep	AFVGTVYRVCLFYIINDGIAHSAPOVRVYLFAPYCGFLPSASDSDLKSSXXSE	234
	orf136ng	AFVGTVYRVCLFYIINDGIAHTAPQVRVYLFAPYCGFLPSASDSDLKSSKYSE	235

The complete length ORF136ng nucleotide sequence <SEQ ID 555> is:

60 1 ATGATGAAGC GCGGTATAGC CGTCTCGTC CTGCTCATGC AGAAAAATCGG
 51 GATTTTGGGA CACTGTTGCG CGAAAAATCGT CACTACAGTT CGGCGACATC

101	GGATGCTCTTG	CCAAATTTTCA	GGGATGATCT	TTTCTCTCAT	ACACCGGCAG
151	TACGTGTCGGG	GAGTGAGCGA	ATGTTTCCTT	CCAGCGCGTA	TGCTGTTCCG
201	TACGTCCTCTC	TTCGCTCATC	TTCGCGGGCA	TTGCGGTGAC	GGTAAAGCCG
251	CGCTAGGGGA	TGCGCTTGCA	CAGCAACATC	CAGTGGTCAG	TGTCGCCAATC
301	CGACGAAACCA	ACGCTTTGCG	CTTGTGACAT	ATTGGTCAGT	CGCGCGGTTT
351	CNTGTTTCAG	CACACAGTCA	ATATAAGGTA	CGTCAAAATA	AATATCGTGG
401	ATTCCAGATAT	ATTGCGAAAT	TTCGCGCTCT	TGCGGCTCTT	GGAAAAGAGG
451	GACTTTGACC	GTTCGAAAT	CCAGGGGCGA	AATAATGCGG	CGCGCTTCCC
501	AAAAAGCTCT	CGCGCAAGAG	TATTTGATTC	TTTATAGGCG	CGCTTCGCGC
551	CGAGCGGTCA	GCGGCTGATC	TTGTTGATTC	TTTATGAAA	CGACGGAATC
601	CGCCATCATC	CATGCTCTCT	CGCTGATGAC	CGCTGTTTGT	CAGTCTCTTG
651	CGCTTTTCTA	CTCTCGGCAT	CCGATATGGA	TTTGAAGATT	TCCAATATT
701	CGGATATG				

This encodes a protein having amino acid sequence <SEO ID 556>:

15 1 MMKRRIVAVF LLMQKIRILG QLLPKIVNTV PAHRMLGQIF GMFFFFIHQR
51 YLPGIAETDS PGGIVFGTLL FTHLSAHCYI GKAAGVDGVA HEHPVADLVN
101 RNNANAFALD IQGSAGCIVFL HVNRIKTKVI NIVDPHFMAN FAVFAVLEKR
151 DHDHGKIQGG NNARAPPKKL APKVFELCTG AFAGTVYRFF CLFYIINDGI
201 AFDHTAPORVR YLFAFYRGFL PVPASDSDLS SKYSE*

20 ORF136ng and ORF136-1 show 93.6% identity in 235 aa overlap:

orf136ng MMKRIIAVFVLLMKIRILGQLLPKIVNVTVAHRMLFIQFGMFFFFIHRQYVLPGIAEIDS
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
orf136-1 MMKRIIAVFVLPQIFRVLGQLLPKIVNVTVAHRMLFIQFGMFFFFIHQQYVLPGIAEIDS

25 orf136ng PGGIVFGTLLFRHLSAHLGYKAAGDVAHEHPVADVNRNANAFALDIGQSAGFIVQ
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
orf136-1 PGGIVFGALLFRHLPAPHLGYKAAGDVAHEHPVADVNRNANAFALDIGQGAGFIVQ

30 orf136ng HTVNIKTVKINIVDPHFMANFAVAVLEKRDFDHKGIOGGNNAAAFPKKLAPKVFCFTGT
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
orf136-1 HTVNIKTVKINIVPFMFANFAVAVLEKRDFDHKGIOGGNNAAAFPKKLAPKFCEFTGT

orf136ng AFAGTVYRFVCLFYINDGAHTHAPOVRVLYFAYRGFLPPASDSDDLSSKYSEX
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
orf136-1 AFVGTVYRFVCLFYINDGAHTHAPOVRVLYFAYRGFLPSADSDDLSSKYSEX

35

Based on the presence of the putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 67

40 The following partial DNA sequence was identified in *N.meningitidis* <SEO ID 557>:

1	GTGCAAAATA	TGCTAACCTT	TTCCAAACAT	AGACCCGATT	TGCGAATCGG
51	CGGCGCGCGG	TTGTTTGGCG	CC_TGCGGAC	GGCGGGAAT	AATGCTGTCCG
101	GGAGCGCGGT	GCAATCAAGC	AAATCCGCGG	CAGTGTGCGG	TTTGGCACTTC
151	CGTGGCGCGG	CATCTAAGAG	TTTGGTCCAT	GTAGGTATTA	TTAAGGTTTCT
201	GAAGGAARAC	GGTATTCTTG	TGAAGGTGTG	TACGCGCACC	TCCGAGGTTT
251	GAATGTGCGG	CACACTTTT	CGATCGGGTA	TCGCGCGGAC	CGCGCTCGAA
301	TTTGGAACAT	GAATATCAG	TGCTGATCAT	TGCTGATCAT	CAGCTGCTGC
351	CCACATGATG	CTATTCAAG	CGCGAAGCT	ATCAGCTGCT	ATCAGCTGCT
401	AATCTCGCGG	CATTCAGATT	CAGCATGATT	CCATCAAAAT	TGCGCGC...

50 This corresponds to the amino acid sequence <SEO ID 558; ORF137>:

```

      1  MENMVTFSKI RPLLAIAAAA LLAAXRTAGN NAVRKPVQTA KPAAVVGLAL
     51  GGGASKGFAH VGIIVLKEN  GIPVKVVTGT SAGSIVGNLF ASGMSFDRLE
    101  LEDELGKTD  LVDITLSTNG EIKGAKLONY INRKIRGMOI QOEPKERA

```

Further work revealed the complete nucleotide sequence <SEO ID 559>:


```

      1 ATGGAAAAATA TGGTAACGTT TTCAAAAATC AGACCGCTTT TGGCAATCGC
    51 CGCGCGCGCG TTGCTTGCGC COTGCGGCAC GCGCGGAAAT AATGCTGCTCC
   101 GCAAGCGCGT GCAAACCGCC AAACCGCGCG CAGTGGTCGG TTTGGCACTC
   151 GGTGGCGCGG CATCTAAAGG ATTTGCCCAT GTAGGTATTA TTAAGGTTTT
   201 GAAAGAAAAAC GGTATTCCCT TGAAGGTGCT TACCGGCACA TCGGCAGGTT
   251 CGATTGTGCG CAGCCTTTTT GCATCGGGTA TGTGCGCCGA CCGCGTCGAA
   301 TTGGAAGCGG AATTTTAGG CAAAACCGAT TTGCTGGATT TAACCTTGTC
   351 CACCACTGGT TTATCAAAG GCGAAAAGCT GCAAAATTAC ATCAACCGAA
   401 AAGTCGCGCG CAGGCAGATT CAGCAGTTTC CCATCAAAAT TCGCGCGCTT
   451 GCTACTGATT TTGAACCGG CAGGCGCGTC GCTTCAATC AGGGAATGTC
   501 CGGCGAGGCT GTGCGCGCTT CGCGCGGCAT TCCCAATGTG TCCCAACCGG
   551 TTATCATCGG CAGGCATACA TATGTTGACG GCGTCTGTC GCAGCCCGTG
   601 CCGCTCAGTG CGCGCCGGCG GCAGGGGGCG AATTTCGTGA TTGCGCTCGA
   651 TATTTCCGCC GCTCCGGGCA AAAACATCAG CCAAGGTTTC TTCTCTTATC
   701 TCGATCAGAC GCTGAACGTA ATGAGCGTTT CTGCGTGTGA AAATGAGTTG
   751 GGGCAGGCGG ATGTGCTTAT CAAACCGCAG GTTTTGGAAT TGGGTGCAGT
   801 CGCGGAGTTC GATCAGAAAA AACGCGCCAT CCGGTTGGGT GAGGAGGCG
   851 CACGTGCGCG ATTGCTGTAA ATCAAAACGA AACTGGCGCG ATACCGTTAT
   901 TGA

```

20 This corresponds to the amino acid sequence <SEQ ID 560; ORF137-1>:

```

      1 MENMVFESKI RFLLAIAAAA LLAACGTAGN NAVRFPVQTA KPAAVVGLAL
    51 GGGASKGFAH VGIILVKLEN GIPVKVVTGT SAGSVGGLLE ASGMSPDRL
   101 LEAEILGKTD LVDLPLSTSG FIKGEKLNQY INRVGGRQI QQPFKFAAV
   151 ATDFETGKAV AFNQGNAGQA VRASAAIPNV FQPIVIGRHT XYVGLGSQLP
   201 PVSAARRQGA NFVIAVDISA RPKNISIQC FSYLDQTLNV MSVSLQNEL
   251 GQADVVIKPO VLDLGVGGF DQKRAIRLG EEAARALPE IKRKLAAERY
   301 *

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

30 ORF137 shows 93.3% identity over a 149aa overlap with an ORF (ORF137a) from strain A of *N. meningitidis*:

```

      10      20      30      40      50      60
orff137.pep MENMVFESKIRPFLLAIAAAALLAAXRTAGNNAVRFPVQTAKPAAVVGLALGGGASKGFAH
      10      20      30      40      50      60
orff137a MENMVFESKIRPFLLAIAAAALLAAGCTAGNNAARFPVQTAKPAAVVGLALGGGASKGFAH
      70      80      90      100     110     120
orff137.pep VGIILVKLENGIPVKVVTGTSAGSVGNLFASGMSPDRLLEAEILGKTDLVDLTLSTNG
      70      80      90      100     110     120
orff137a VGIILVKLENGIPVKVVTGTSAGSVGSLFASGMSPDRLLEAEILGKTDLVDLTLSTSG
      130     140     149
orff137.pep FIKGAKLQNYINRKLGRMQIQQPFKFAA
      130     140     150     160     170     180
orff137a FIKGEKLNQYINRKLGRRIQQPFKIFAAVATDFETGKAVAFNQGNAGQAVRASAAIPNV

```

The complete length ORF137a nucleotide sequence <SEQ ID 561> is:

```

      1 ATGGAAAAATA TGGTAACGTT TTCAAAAATC AGACCGCTTT TGGCAATCGC
    51 CGCGCGCGCG TTGCTTGCGC COTGCGGCAC GCGCGGAAAT AATGCTGCTCC
   101 GCAAGCGCGT GCAAACCGCC AAACCGCGCG CAGTGGTCGG TTTGGCACTC
   151 GGTGGCGCGG CATCTAAAGG ATTTGCCCAT GTAGGTATTA TTAAGGTTTT
   201 GAAAGAAAAAC GGTATTCCCT TGAAGGTGCT TACCGGCACA TCGGCAGGTT
   251 CGATAGTCGG CAGCCTTTTT GCATCGGGTA TGTGCGCCGA CCGCGTCGAA
   301 TTGGAAGCGG AATTTTAGG TAAACCGAT TGTGCTGGAT TAACCTTGTC
   351 CACCACTGGT TTATCAAAG GCGAAAAGCT GCAAAATTAC ATCAACCGAA
   401 AAGTCGCGCG CAGGCAGATT CAGCAGTTTC CCATCAAAAT TCGCGCGCTT
   451 GCTACTGATT TTGAACCGG CAGGCGCGTC GCTTCAATC AAGGGAATGC
   501 CGGCGAGGCT GTGCGCGCTT CGCGCGGCAT TCCCAATGTG TCCCAACCGG
   551 TTATCATCGG CAGGCATACA TATGTTGACG GCGTCTGTC GCAGCCCGTG

```

5
 601 CCCGTCAGTS CCGCCCGGCG GCANGNNNG NATNTCGTGA TTGCGGTGGA
 651 TATTTCGCCG CGTCCGAGCA AAAACATCAG CCAAGCGCTC TTCTCTTATC
 701 TCGATCAGAC GCTGAACGTA ATGAGCGGTT CCGCGTTGCA AATGAGTTG
 751 GGGCAGCGCG ATGTGGTTAT CAAACGSCAG GTTTTGGATT TGGGTGCGAT
 801 CGCGGATTG GATCAGAAAA AACCGCCCAT CCGGTGGGT GAGGAGCGAG
 851 CACGTGCCGC ATTGCCTGAA ATCAAACGCA AACTGGCGCG ATACCGTTAT
 901 TGA

This encodes a protein having amino acid sequence <SEQ ID 562>:

10
 1 MENMVFESKI RPLAIAAAAA LLAACGTAGN NAARKPVQTA KPAAVVGLAL
 51 GCGASKGFAH VGIITKVLKN GIPVKVVTGT SAGSIVGSLF ASGMSPDRIE
 101 LEAEILGKTD LVDLTLSTSG FIKGEKLQNY INRKVGRRRI QQFPFKFAAV
 151 ATDFETGKAV AFNQGNAQGA VRASAAIPNV FQPVIIGRHT YVDGGLSQPV
 201 PVSAARRXXX XXVIAVDISA RPSKNISQGF FSYLDQTLNV MSVSALQNEL
 251 GQADVVIKQP VLDLGAVGGF DQKKRAIRLG EEAARAALPE IKRKLAAVRY
 15
 301 *

ORF137a and ORF137-1 show 97.3% identity in 300 aa overlap:

20
 orf137a.pep MENMVFESKIRPLAIAAAAA LLAACGTAGNNAARKPVQTA KPAAVVGLALGGGASKGFAH
 orf137-1 MENMVFESKIRPLAIAAAAA LLAACGTAGNNAVRKPVQTA KPAAVVGLALGGGASKGFAH
 orf137a.pep VGIITKVLKENGIPVKVVTGT SAGSIVGSLFASGMSPDRIELEAEILGKTDLVDLTLSTSG
 orf137-1 VGIITKVLKENGIPVKVVTGT SAGSIVGSLFASGMSPDRIELEAEILGKTDLVDLTLSTSG
 25
 orf137a.pep FIKGEKLQNYINRKVGRRIQQFPFKFAAVATDFETGKAVAFNQGNAQGA VRASAAIPNV
 orf137-1 FIKGEKLQNYINRKVGRRIQQFPFKFAAVATDFETGKAVAFNQGNAQGA VRASAAIPNV
 30
 orf137a.pep FQPVIIGRHTYVDGGLSQPVPSAARRXXXVIAVDISARPSKNISQGF FSYLDQTLNV
 orf137-1 FQPVIIGRHTYVDGGLSQPVPSAARRQGANFVIAVDISARPGKNI SQGF FSYLDQTLNV
 orf137a.pep MSVSALQNELGQADVVIKQP VLDLGAVGGFDQKKRAIRLGEEAARAALPEIKRKLAAVRY
 35
 orf137-1 MSVSALQNELGQADVVIKQP VLDLGAVGGFDQKKRAIRLGEEAARAALPEIKRKLAAVRY

Homology with a predicted ORF from *N.gonorrhoeae*

ORF137 shows 89.9% identity over a 149aa overlap with a predicted ORF (ORF137ng) from

N.gonorrhoeae:

40
 orf137.pep MENMVFESKIRPLAIAAAAA LLAAXRTAGNNAVRKPVQTA KPAAVVGLALGGGASKGFAH 60
 orf137ng MENMVFESKIRSLAIAAAAA LLAACGTAGNNAARKPVQTA KPAAVVGLALGGGASKGFAH 60
 45
 orf137.pep VGIITKVLKENGIPVKVVTGT SAGSIVGNL FASGMSPDRIELEAEILGKTDLVDLTLSTNG 120
 orf137ng IGIVKVLKENGIPVKVVTGT SAGSIVGSL LAGMSPDRIELEAEILGKTDLVDLTLSTNG 120
 orf137.pep FIKGAKLQNYINRKLGRMQIQFPFKFAA 149
 50
 orf137ng FIKGEKLQNYINRKVGRRIQQFPFKFAAVATDFETGKAVAFNQGNAQGA VRASAAIPNV 180

The complete length ORF137ng nucleotide sequence <SEQ ID 563> is:

1 ATGGAATA TGATAACGTT TTCAAAATC AGATCATTTT TGGCAATCGC
 51 GCGCGCGCGG TTGCTTGCGG CCGGCGGTAC GCGGGAAAC AATGCGCGCC
 101 GCAAGCGCGG GCAACCGCGC AACCGCGCGG CAGTGGTCGG TTGCGCACTC
 151 GGTGCGCGCG CATCTAAAGG ATTTGCCCAT ATGGAATTG TTAAGGTTTT
 201 GAAGAAACG GTATTCTCGT TGAAGTGGT TACCGGACA TCGGCGGTT
 251 CAGTACTCGG CAGCTCTTGG GCATCGGGA TGTGCGCGA CCGCTCGAA
 301 TTGGAACGCG AGATTTTAGG TAAACCGAT TTAGTCGATT TAACCTTGT
 351 CACCACTGGT TTTATCAAGG GCGAAAGCT GCAAAATTAC ATCAACCGAA
 60
 401 AAGTGCAGCG CAGGCAGATT CAGCAGTTTC CCATCAAAAT TGCGCGCGTT
 451 GCCACTGATT TTGAAACCGG CAAGGCGCTC GCTTTCAATC AAGGGAATGC

501 CGGGCAGGCG GTTCGTGCTT CCGCGCCAT TCCCAATGTG TTCCAGCCAG
 551 TCATCATCGG CAGGCACAAA TATGTTGACG GCGTCTGTGC GCAGCCCGTG
 601 CCGCTCAGTG CCGCTCGGCG CAGGGGGCG AATTTCGTGA TTGCGCTGCA
 651 TATTTCCGCA CGTCCGAGCA AAAATGTCGG TCAAGGTTTC TTCTCTTATC
 701 TCGATCAGAC GCTGAACGTG ATGAGCGTTT CCGTGTTCGA AAACGAGTTG
 751 gggcAGGCGG ATGTGGTTAT CAACCCGCG gtTTTGATT TGGTGCAAT
 801 CCGCGGATTC GATCAGAAAA AGCGCCCAT CCGTTGGGC GAGGAGGACG
 851 CAGTCCGCGC ATTGCCTGAA ATCAACGCA AACTGCGGCG ATACCGTTAT
 901 TGA

10 This encodes a protein having amino acid sequence <SEQ ID 564>:

1 MENMVFPSKI RSFLAIAAAA LLAACGTAGN NAARKPVQTA KPAAVVALAL
 51 GGGASKGFAH IGIVKLVKEN GIPVKVVTGT SAGSIVGSL L ASGMSPDRL
 101 LEAEILGKTD LVDLTLTSTG FIKGEKLQNY INRKVGGRI QQFPIKFAV
 151 ATDFETGKAV AFNQGAGQA VRASAIIPNV FQPVIIGRHK YVDGGLSQPV
 201 PVSAARRQGA NFVIAVDISA RPSKNVGGGF FSYLDQTLNV MSVSVLQNEL
 251 GQADVVIKPO VLDLGA VGGF DQKKRAIRLG EEAARALPE IKRKLAAARY
 301 *

ORF137ng and ORF137-1 show 96.0% identity in 300 aa overlap:

20 orf137ng MENMVFPSKIRSFLLAIAAAALLAACGTAGNNAARKPVQTA KPAAVVALALGGGASKGFAH
 orf137-1 MENMVFPSKIRFLAIAAAALLAACGTAGNNAVRKPVQTA KPAAVVGLALGGGASKGFAH
 25 orf137ng IGIVKLVKENGIPVKVVTGTSAGSIVGSLASGMSPDRLLEAEILGKTDLVDLTLTSTG
 orf137-1 VGIIKLVKENGIPVKVVTGTSAGSIVGSLFASGMSPDRLLEAEILGKTDLVDLTLTSTG
 30 orf137ng FIKGEKLQNYINRKVGGRIQQFPIKFAAVATDFETGKAVAFNQGAGQAVRASAIIPNV
 orf137-1 FIKGEKLQNYINRKVGGRIQQFPIKFAAVATDFETGKAVAFNQGAGQAVRASAIIPNV
 35 orf137ng FQPVIIGRHKYVDGGLSQPVVPSAARRQGANFVIAVDISARPSKNVGGGFYSYLDQTLNV
 orf137-1 FQPVIIGRHTYVDGGLSQPVVPSAARRQGANFVIAVDISARPKNISQGFYSYLDQTLNV
 40 orf137ng MSVSVLQNELGQADVVIKPVLDLGA VGGF DQKKRAIRLGEEAARALPEIKRKLAAARY
 orf137 MSVSALQNELGQADVVIKPVLDLGA VGGF DQKKRAIRLGEEAARALPEIKRKLAAARY

Based on the presence of a predicted prokaryotic membrane lipoprotein lipid attachment site (underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 68

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 565>:

45 1 ATGTTTCGTT TACAATTCAG GCTGTTTCCC CTTTGCAGAA CCGCATGCA
 51 CATCCTGTTG ACCGCCCTGC TCAAATGCTC CTCCCTGTGT CCGCTTCTCT
 101 GTCTGCACAC GCTGGGAAC CCGCTCGGAC ATCTGGCGTT TTACCTTTTA
 151 AAGGAAGACC GCGCGGCAT CGTGCGChAT ATGCGCGAGG CGGCTTGA
 201 CCGCGACCCC AAAACGGTCA AAGCGGTTTT TGCGGAACG GCAGAAAGGCG
 251 GTTTGGAACT TCGCCCGCG TTTTTCAGAA AACCGGAAGA CATAGAAACA
 50 ATGTTCAAG CCGTACCGG CTGGGACAT GTGCAGAGG CTTTGACAA
 351 ACACGAAGG CTGCTATTC . .

This corresponds to the amino acid sequence <SEQ ID 566; ORF138>:

1 MFRQLRFLFP PLRTAMHILL TALLKCLSL PLSCLHTLGN RLGHIAFYLL
 51 KEDRARIVAX MRQAGLNPD KTVKAVFRET AKGLLELAPA FFRKPEDIE
 101 MFAVHGWEH VQALDKHEG LLF

Further work revealed the complete nucleotide sequence <SEQ ID 567>:

```

1 ATGTTTGGTT TACAATTCAG GCTGTTTCCC CTTTGGGAA CGGCCATGCA
51 CATCCTGTTG ACCGCGCTGC TCARATGCTT CTTCCCTGCTG CCGCTTTTCT
101 GTCTGCAAC GCTGGGAAC CGCTCGGAC ATCTCGGCTT TTACCTTTTA
151 AAGGAAGACC CGCGCGCAT COTCGCCAA ATCGCGCAGG CCGGTTTGAA
201 CCCCCACCCC AAAACGGTCA AAGCGGTTT TCGCGAAGG GCAAAGGCG
251 GTTTGGAAC TGCCTCCGCG TTTTTCAGAA AACCGAAGA CATAGAAACA
301 ATGTTCAAAG CGGTACACGG GTGGGAACAT GTGCAGCAGG CTTTGACAAA
351 ACACGAAGGG CTGCTATTCA TCACGCGCA CATCGGCAGG TACGATTGG
401 GCGGACGCTA CATCAGCCAG CAGCTTCGT TCCCGCTGAC CGCCATGTAC
451 AAACCGCGCA AATCAAAGC GATAGACAAA ATCATGCAGG CGGGCAGGTT
501 TCGCGGCAAA GAAAAACCG CGCTACGAG CATACAAGG GTCAACACAA
551 TCATCAAAGC CTTGCGTTCC GCGGAAGCAA CATTGCTCT CCGCGACCAC
601 GTCCCTCTCC CTCAGAAGG CGGGGAAGGC GTATGGTGG ATTCTTCGG
15 651 CAAGCTGCC TATACATGA CGCTGGCGG AAAATTGGCA CAGCTCAAG
701 GCGTGAAGC CTTGTTTTC TCTGCGAAC CCGTGCCTGG CGGCAAGGT
751 TTCGATTGGC ACATCCGCC CCGCAAGGG GAATTGAAG GCGCAAGC
801 CCAATGATGC GCGGTGTCA ACCGCAATG CAAATTATG ATACGCCGTT
851 TTCCGACGCA GTATCTGTTT ATGTACAACC GCTACAATAT CGCGTAA

```

20 This corresponds to the amino acid sequence <SEQ ID 568; ORF138-1>:

```

1 MFRLQRLFP PLRTAMHILL TALLKCLSL PLSCLHTLGN RLGLHAFYLL
51 KEDRARIVAN MRQAGLNPDPT KTVKAVFAET AKGGLLELAPA FFRKPEDTET
101 MFKAVHGWEN VQALDKHEG LFLITPHIG YDLGGYISQ QLPFPLTAMY
25 151 KPPKIKAIKD IMQAGRVKRG KTAFTPTSIQ VQKIKALRS GEATVILPDD
201 VPSQEGGEG VWVDFEKGPA YMTLAAKLA HVKGKTLFF CCELRPGQGG
251 FDLHIRPVQG ELNGDKAADA AVFNRAEYN TRFPTQYLF MYNRYKMP*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF138 shows 99.2% identity over a 123aa overlap with an ORF (ORF138a) from strain A of *N.*

30 *meningitidis*:

```

              10      20      30      40      50      60
orf138.pep MFRLQRLFPFLRTAMHILLTALLKCLSLPLSCLHTLGNRLGLHAFYLLKEDRARIVAX
              |||
orf138a MFRLQRLFPFLRTAMHILLTALLKCLSLPLSCLHTLGNRLGLHAFYLLKEDRARIVAN
35              10      20      30      40      50      60
              70      80      90      100     110     120
orf138.pep MRQAGLNPDPTKTVKAVFAETAKGGLLELAPAFFRKPEDTETMFKAVHGWENVOALDKHEG
              |||
orf138a MRQAGLNPDPTKTVKAVFAETAKGGLLELAPAFFRKPEDTETMFKAVHGWENVOALDKHEG
40              70      80      90      100     110     120
orf138.pep LLF
              |||
orf138a LLFITTPHIGSYDLGGYISQQLFPPLTAMYPKPKIKAIKIMQAGRVKRGKRTAFTPTSIQ
              130     140     150     160     170     180

```

The complete length ORF138a nucleotide sequence <SEQ ID 569> is:

```

1 ATGTTTGGTT TACAATTCAG GCTGTTTCCC CTTTGGGAA CGGCCATGCA
50 51 CATCCTGTTG ACCGCGCTGC TCARATGCTT CTTCCCTGCTG CCGCTTTTCT
101 GTCTGCAAC GCTGGGAAC CGCTCGGAC ATCTCGGCTT TTACCTTTTA
151 AAGGAAGACC CGCGCGCAT COTCGCCAA ATCGCGCAGG CCGGTTTGAA
201 TCGGACCCC AAAACGGTCA AAGCGGTTT TCGCGAAGG GCAAAGGCG
55 251 GTTTGGAAC TGCCTCCGCG TTTTTCAGAA AACCGAAGA CATAGAAACA
301 ATGTTCAAAG CGGTACACGG GTGGGAACAT GTGCAGCAGG CTTTGACAAA
351 AGACGAAGGG CTGCTATTCA TGAGGGGAGA CATCGGCAGG TACGATTGG
401 GCGGACCGTA CATCAGCCAG CAGCTTGGGT TCCCGCTGAC CGCCATGTAC
451 AAACCGGCGA AATCAAAGC GATAGACAAA ATCATGCAGG CGGGCAGGTT
60 501 TCGCGGCAAA GAAAAACCG CGCTACGAG CATACAAGG GTCAACACAA
551 TCATCAAAGC CTTGCGTTCC GCGGAAGGAA GCATGCTCT CCGCGACCAC

```

5 601 GTCCCTCCCT CTCAAGAAGG CGGGGAAGGC GTATGGGTGG ATTTCTTCGG
 651 CAAACCTGCC TATACCATGA CGCTGGCGGC AAAATTGGCA CACGTCAAAG
 701 GCGTGAAGAC CCGTGTCTTC TGCTGGGAAC GCGTGCCTGG CGGACAAGGT
 751 TTGATTTTGG ACATCCGCCCG CGTCCAAGGG GAATTGAACG GCGACAAGCG
 801 CCAATGATGCC GCGGTGTTC ACGGCAATGC CGAATTATTG ATACGCGGTT
 851 TTCGACGCGA GTATCTGTTT ATGTACAACC GCTACAAAT GCGGTAA

This encodes a protein having amino acid sequence <SEQ ID 570>:

10 1 MFRQLFRLFP PLRTAMHILL TALLKCLSL PLSCHLTLGN RLGHAFAYLL
 51 KEDRARIIVAN MRQAGLNPDPT KTKAVFAET AKGGLLELAPA FFRKPEDJET
 101 MFKAVHWEH VQQALDKHEG LLFTPHIGS YDLGRVYISQ QLPFFLTAMY
 151 KFPKIKALDK IMQAGVRVKG GKTAPTSIQG VKQIKALRS GEATIVLPDH
 201 VPSPPQEGGEG VVWDFGKFA YMTLAAKLA HVKGVKTLFF CCERLPGGGG
 251 FDLHIRPVQG ELNGDKAHDA AVFNRRNAEYW IRRFPTQYLF MYNRYKMP*

ORF138a and ORF138-1 show 99.7% identity over a 298aa overlap:

15 orf138a.pep MFRQLFRLFPPLRTAMHILLTALLKCLSLPLSCHLTLGNRLGHAFAYLLKEDRARIIVAN
 orf138-1 MFRQLFRLFPPLRTAMHILLTALLKCLSLPLSCHLTLGNRLGHAFAYLLKEDRARIIVAN
 20 orf138a.pep MRQAGLNPDPTKTKAVFAETAKGGLLELAPAFRRKPEDJETMFKAVHWEHVQQALDKHEG
 orf138-1 MRQAGLNPDPTKTKAVFAETAKGGLLELAPAFRRKPEDJETMFKAVHWEHVQQALDKHEG
 25 orf138a.pep LLFITPHIGSYDLGGYISQQLPFPLTAMYPKPKIKAKIDKIMQAGVRVKGKGTAPTSIQG
 orf138-1 LLFITPHIGSYDLGGYISQQLPFPLTAMYPKPKIKAKIDKIMQAGVRVKGKGTAPTSIQG
 30 orf138a.pep VKQIKALRSGEATIVLPDHVPSPPQEGGEGVWVDFGKPAYTMTLAAKLAHVKGVKTLFF
 orf138-1 VKQIKALRSGEATIVLPDHVPSPPQEGGEGVWVDFGKPAYTMTLAAKLAHVKGVKTLFF
 35 orf138a.pep CCERLPGGQGFDLHIRPVQGEELNGDKAHDAAVFNRRNAEYIIRRFPTQYLFMYNRYKMP
 orf138-1 CCERLPGGQGFDLHIRPVQGEELNGDKAHDAAVFNRRNAEYIIRRFPTQYLFMYNRYKMP

35 Homology with a predicted ORF from *N.gonorrhoeae*

ORF138 shows 94.3% identity over a 123aa overlap with a predicted ORF (ORF138ng) from *N.gonorrhoeae*:

40 orf138.pep MFRQLFRLFPPLRTAMHILLTALLKCLSLPLSCHLTLGNRLGHAFAYLLKEDRARIIVAX 60
 orf138ng MFRQLFRLFPPLRTAMHILLTALLKCLSLPLSCHLTLGNRLGHAFAYLLKEDRARIIVAN 60
 orf138.pep MRQAGLNPDPTKTKAVFAETAKGGLLELAPAFRRKPEDJETMFKAVHWEHVQQALDKHEG 120
 orf138ng MRQAGLNPDPTKTKAVFAETAKGGLLELAPAFRRKPEDJETMFKAVHWEHVQQALDKHEG 120
 45 orf138.pep LLF 123
 orf138ng LLFITPHIGSYDLGGYISQQLPFPLTAMYPKPKIKAKIDKIMQAGVRVKGKGTAPTSIQG 180

The complete length ORF138ng nucleotide sequence <SEQ ID 571> is:

50 1 ATGTTTCGTT TACAATTCAG GCTGTTTCCC CTTTGGCGAA CGGCCATGCA
 51 CATCCTGTGT ACCGCCCTGC TCAAAATGCCT CTCCTGTCTG TCGCTTTCCT
 101 GTCTGCACAC GCTGGGAACG CGGCTCGGAC ATCTGGCGTT TTAACCTTTA
 151 AAGGAAGACC GCGCGCGCAT CGTGGCCAAT ATGGCGCAGG CGGGTTTGAA
 201 CCGCGACACG CAGACGGTCA AAGCGGTTTT TGGCGAAGCG GCAGAAATGG
 55 251 GTTTGGAACT TGGCCCGCGG TTTTCAAAA AACCGGAAGA CATCGAAACA
 301 ATGTTCAAGG CGGTACACGG CTGGGAACAC GTGCAGCAGG CTTTGGACAA
 351 GCGCGAAGCG CTGCTGTTC TACGAGCGCA CAGCGCAGC TACGATTTGG
 401 GCGAGCGCTA CATGACGCCG CAGCTTCGCT TCAACCTGAC CGGCATGTAC
 451 AAGCGCGCGA AATCAAGCG GATAGACAAA ATCATCGAGG CGGGCAGGCT
 60 501 GCGCGGCRAA GCGAAAACcg cgccccagcg catACAAGGG GTCAACACAA
 551 tcatcaAGGC CCGCGCGCGG GCGGAGGCAA CCAATCTCT GCGCGACCAC

5
 601 GTCCTCTCTC CGCAGGAagg cggCGGCGTG TGCGCGGATT TTTTCGGCAA
 651 ACCTGCATAC acCATGACAC TGCGCGGCAA ATTGCGACAC GTCAAAGGCG
 701 TGAAGAACCT GTTTTCTGCG TGCGAACGCC TGCCCGACGG ACAAGGCTTC
 751 GTGTGTCACA TCCGCCCCGT CCAAGGGGAA TTGAACGCGCA ACAAGGCCCA
 801 CGATGCGCGC GTGTTCACCC GCAATACCGA ATATTGGATA CGCGTTTTC
 851 CGACGCGAGTA TCTGTTTATG TACAACCGCT ATAAACGCC GTAA

This encodes a protein having amino acid sequence <SEQ ID 572>:

10
 1 NFRQLRFLRP PLRTAMHILL TALLKCLSL SLSCSLHTLGN RLGHIAFYLL
 51 KEDRARIVAN MRQAGLNPDQT VKVFAETAKGGLLELAPAFRRKPEDINETMFKAVHGVHVVQALDKHEG
 101 MFKAVHGVHVVQALDKHEG LLFTPHIGS YDLGRYISQQLPFLHTAMY
 151 KPPKIKAIKD IMQAGVRVGK GKTAPTGIQGV KQIIKALRA GEATILPDH
 201 VPSPOEGGVWADFFGKPAY TMTLAALKLAH VKGVKTLFFC CERLPDGGGF
 251 VLHIRPVQGE LNGNKAHDA VFNRNTEYWI RRPFTQYLFM YNRKYTP*

ORF138ng and ORF138-1 show 94.3% identity over 299aa overlap:

15
 orf138-1.pep NFRQLRFLRPPLRTAMHILLTALLKCLSLSLSCSLHTLGNRLGHIAFYLLKEDRARIVAN
 orf138ng NFRQLRFLRPPLRTAMHILLTALLKCLSLSLSCSLHTLGNRLGHIAFYLLKEDRARIVAN
 20
 orf138-1.pep MRQAGLNPDQTKVAVFAETAKGGLLELAPAFRRKPEDINETMFKAVHGVHVVQALDKHEG
 orf138ng MRQAGLNPDQTKVAVFAETAKGGLLELAPAFRRKPEDINETMFKAVHGVHVVQALDKHEG
 25
 orf138-1.pep LLFTPHIGSYDLGGRYISQQLPFLHTAMYKPPKIKAIKD KIMQAGVRVGKGTAPTGIQGV
 orf138ng LLFTPHIGSYDLGGRYISQQLPFLHTAMYKPPKIKAIKD KIMQAGVRVGKGTAPTGIQGV
 30
 orf138-1.pep VKQIIKALRAGEATILPDHVPSPQEGGEGVWVDFGKPAYTMTLAALKLAHVKGVKTLFF
 orf138ng VKQIIKALRAGEATILPDHVPSPQEGGEGVWVDFGKPAYTMTLAALKLAHVKGVKTLFF
 35
 orf138-1.pep CCERLPDGGGFVLHIRPVQGE LNKGDKAHDAAVFNRAEYWRIRFPFTQYLFMYNRYKMP
 orf138ng CCERLPDGGGFVLHIRPVQGE LNKGDKAHDAAVFNRAEYWRIRFPFTQYLFMYNRYKMP

In addition, ORF138ng is homologous to htrB protein from *Pseudomonas fluorescens*:

35
 gnl:PID|e334203 (Y14568) htrB [Pseudomonas fluorescens] Length = 253
 Score = 80.8 bits (196), Expect = 9e-15
 Identities = 49/151 (32%), Positives = 79/151 (51%), Gaps = 6/151 (3%)
 40
 Query: 101 MFKAVHGVHVVQALDKHEG LLFTPHIGSYD-LGGRYISQQLPFLHTAMYKPPKIKAIKD 159
 +++ V G E ++AL G+G+ IT H+G+++ L Y S Q P Y+PK+KA+D
 Sbjct: 94 LVREVEGLEVLKEALASGKGVGITSHLGNWEVLNHFYCSQCKPI---FYRPPKLAVD 150
 45
 Query: 160 KIMQAGVRVGKGTAPTGIQGVKQIIKALRAGEATILPDHVPSPQEGGEGVWVDFGKPA 219
 ++++ RV+ KA+ +G+ +IK+R G I D P P E G++ FF A
 Sbjct: 151 ELLRKQVRQLGNKVAASKEGILSVIKEVRKGGQVGPAD--PEAESAGIFVPFFATQA 208
 Query: 220 YMTLAALKLAHVKGVKTLFFCCERLPDGGGF 250
 T + +F RLPDG G+
 Sbjct: 209 LFSKFVPMMLAGGKAVGVFLHALRLPDGGGY 239

50 Based on this analysis, including the presence of a putative transmembrane domain in the gonococcal protein, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF138-1 (57kDa) was cloned in the pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 14A shows the results of affinity purification of the GST-fusion protein. Purified GST-fusion protein was used to immunise mice, whose sera were used for ELISA (positive result) and FACS analysis

(Figure 14B). These experiments confirm that ORF138-1 is a surface-exposed protein, and that it is a useful immunogen.

Example 69

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 573>:

```

5       1  ..GCGTGGTCGG  CGGGCGAATC  GTGGCGTGTG  TTAATGGAAA  GTGAAACGTG
51      GCATGCGGCTG  TGGAACTACT  TGCCTTCTCT  GCGCGCGGCG  GTGTATGCGG
101     CAGCGGTTTTT  GGGTGTGGTG  TATGCGGCGC  CGGCGCGGCG  GTGCGCGTTC
151     ATGCGCGGGC  TGAATGTTTA  GCCGTTATAT  GTGTGCGCGG  TTGTGTTTTC
201     GCGCGGCGTG  CTGCTGCTTT  ATCCGCAATG  GACGGCTTCG  TTGCGCTTGC
10      251     TGCTGGCGGAT  GTATGCGCTG  CTGGCGTATC  CGTTTGTGGC  AAAAGATGTT
251     TTATCAGCCT  GGGATGCACT  GCCCGGATG  TACGCAAGGG  CGGCGCGCGG
351     TTTGGGTGCA  AACGCGTTTC  AGACGATGCG  CCGCATCAGC  TTCCCGCTGT
401     TGAACCGGCG  GTTGCAGCGC  GGTCTGACTT  TGGCGCGGCG  AACCTGCTGT
451     GCGGAATTGT  CGCGGACATT  GTTTCGTGTC  CGTCCGGAAT  GGCAGACGCT
15      501     GACGACTTTG  ATTATGCTCT  ATTTGGGACG  CCGGGGTGAG  GATAATTACG
551     CGCGGGCGAT  GGTGCTG...

```

This corresponds to the amino acid sequence <SEQ ID 574; ORF139>:

```

1       1  ..AWSAGESWRV  LMESETHAV  WNTLRFSAAA  VYAAAVLGVV  YAAPARRSAW
51      MRGLMFPEFM  VSPVCSAGV  LLLYPWTAS  LPLLAMYAL  LAYPFVAKEV
20      101     LSANDALFPD  YGRAAAGLA  NGFQACRIT  FPLLKPAIRR  GLTAAATCV
151     GEFARTLFLS  RPEWQTLTTL  IYAYLGRAGE  DNYARAMVL...

```

Further work revealed the complete nucleotide sequence <SEQ ID 575>:

```

1       1  ATGGATGAGAC  GCGGTTGGGT  GGTATGGGGT  GCTTTTGCCC  TGCTGCCCTTC
51      GGCCTTTTTT  GCGGTAATGG  TCGTTGGGCG  TTTGTGGGCG  GTGGCGGCGT
25      101     ATGACGGTTT  GCGGTGGGCG  CGGCTGCTGT  CGGATGCCTA  TATGCTCAAA
151     CGTTTGGCGT  GGAACGGTAT  TCAAGCAGCG  GCAACCTGTG  TGCTGGTGTCT
201     GCCTTTGGGC  GTGCGTGTGG  CGTGGGTGCT  GCGCGGCGTG  GCGTTTCGGG
251     GCGCGGCGTT  GGTGCTGGCG  CTGCTGATGC  TGCCCTTTGT  GATGCCCCAG
301     TTTGGTGGCG  GCGTGGGCGT  GCTGGGCCGT  TTCGGGCGCG  ACGGGCTGTT
30      351     GTGGCGGCGC  AGGCAGGATA  CGCCGTATCT  GTTGTGTGAC  GGCATGTGTT
401     TTTTCAACCT  TCCTGTGTTG  GTCAGGGGCG  CGTATCAGGG  GTTGTGCAAA
451     GTGCTGCGCG  CACGGCTTCA  GACGGCAGCG  ACGTTGGGCG  CGGGGGCGTG
501     GCGCGGCTTT  TGGGACATTG  AAATGCCCGT  TTTGCGCCCG  TGCGTTGCCG
35      551     GCGGCTGTGT  CTTTGTCTTT  CTGTATTGTT  TTTCCGGGTT  CGGGCTGGCG
601     CTGCTGCTGG  GCGGCAAGCG  TTATGCCACG  GTGAAATGG  AAAATTACCA
651     GTTGGTCAAT  TTGAACAATG  ATATGGCGGT  TGCTTGGGTT  CTGGTGTGGC
701     TGGTGTGGGG  GGTAAACGGG  CGCGCAGGGT  TGCCTGATGC  GTGCTTCGGC
40      751     AGGCGCGCGG  TTTTGGGATA  GCGCGTTTTC  CCTGTGATGC  GTGCGCGCGC
801     CAGTTCGGCT  GGGGAAATAG  TCGTGTGGCG  GTTTCGGGCG  GCGGTGTGTT
851     CTGTGTGGCG  TGTGTGGCAA  TTGTGTGGAA  TTGTGTGGAA  AGCGTGTGCG
901     GCGCGCAAT  CGTGGCGTGT  GTTAAATGAA  AGTGAACGCT  GCGCAGCGGT
951     GTTGAATACT  TTGCGCTTCT  CGCGCGCGGC  GGTGTATGCG  GCGCGGTTTT
1001    TGGGTGTGGT  GTATGCGGCG  GCGCGCGCGC  GGTGCGCGTG  GATGCGCGGG
1051    CTGATGTTTT  TGCGGTTTAT  GGTGTGCGCG  GTTGTGTTT  CGCGGGCGGT
45      1101    GCTGCTGCTT  TATCGCGAGT  GGACGGCTTC  GTTGGCGGTG  CTGCTGGCGA
1151    TGTATGCGCT  GCTGGGCGAT  CGCTTTGTGG  CAAAAGATGT  TTATTCAGCC
1201    TGGGATGCAAT  TGCGCGCGGA  TTACGGCAGG  GCGCGCGCGG  GTTGGGTGTC
1251    AAAACGGCTT  CAGACGGCAT  CGCGCATCAC  GTTCCCTCTC  TTGAACCGGG
1301    CGTTGCGGCG  CGGTCTGACT  TTGGCGGCGG  CAACCTGCGT  GGGCGAATTT
50      1351    GCGCGGCAAT  TGTTTCTGTC  CGCTCCGGAA  TGGCAGACGC  TGACGACTTT
1401    GATTATGCTC  TATTGCGGAC  GCGCGGGTGA  GGATATATTC  GCGCGGCGCA
1451    TGGTGTGCTG  ATTGCTGTTG  GCGCGCTTGC  CGCTGGGTAT  TTTCTGCTGT
1501    TTGACGCGCG  GCGAAGCGCG  AAAACGAGCG  GAAAGCTTAT  AA

```

This corresponds to the amino acid sequence <SEQ ID 576; ORF139-1>:

```

55      1  MDGRRWVVMG  AFALLPSAFL  AVMVVAPLNA  VAAVYDGLAW  AVLSDAYMLK
51      RLAWTVFQAA  ATCVLYLPLG  VPVAVWLARI  AFPRGALVLR  LLMFLPFVMT
101     LVAGVGVIAL  FGADGLLWRG  RQDTPYLLLY  GNVEFNLPVL  VRAAYQCFVQ
151     VFAARLQIAR  TLGAGAWRRF  NDIEMPVLRP  WLAGGVCLVF  LYCFSGFGLA

```

-330-

201 LLLGGSRVAT VEVETIYQLVM FELDMAVASV LVWLVLGVTA AAGLLYAWFG
 251 RRAVSDKAVS PUMSPPPQSV GEYVLLAFAP AVLSVCCLFP LLAIVVKAMS
 301 AGESWRVLMSE SETQVAVWNT LRFSAAVYA AAVLGUVYAA AARRSAMMRG
 351 LMFLLPMVSP VCVSAGVLLL YPQWATSLPL LLAMYALLAY PFVAKDVLSA
 401 WDALPPDYGR AAGLGGANGF QTACRITFPL LKFAIRRGIT LAAATCVGEF
 451 AATLFLSRPE WQTLTTLIYA YLGRAGEDNY ARAMVLTLLL AAFALGIFLL
 501 LDGEGGKQT ETL*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

10 ORF139 shows 94.7% identity over a 189aa overlap with an ORF (ORF139a) from strain A of *N. meningitidis*:

```

      10      20      30
orf139.pep      AWSAGESWRVLMSESTWHAVWNTLRFSAAA
15 orf139a      QSVGEYVLLAFAPAAVSVCCFLXLLAIVVKAMWAGESWRVLMSESTWQAVWNTXRFSA
      270      280      290      300      310      320

      40      50      60      70      80      90
orf139.pep      VYAAAVLGUVYAAPARRSAMMRGLMFXPFMVSPVCVSAGVLLLYPQWATSLP
20 orf139a      VYAAAVLGUVYAAARRSAMMRGLMFLPMVSPVCVSAGVLLLYPQWATSLP
      330      340      350      360      370      380

      100      110      120      130      140      150
orf139.pep      LAYPFVAKDVLSDALPDDYGRAAAGLGANGFQTACRITFPLLPKALRRGLT
25 orf139a      LAYPFVAKDVLSDALPDDYGRAAAGLGANGFQTACRITFPLLPKALRRGLT
      390      400      410      420      430      440

      160      170      180      189
orf139.pep      GEFAATLFLSRPEWQTLTTLIYAYLGRAGEDNYARAMVL
30 orf139a      GEFAATLFLSRPEWQTLTTLIYAYLGRAGEDNYARAMVL
      450      460      470      480      490      500

```

35 The complete length ORF139a nucleotide sequence <SEQ ID 577> is:

```

1  ATGGATGGAC  GCGCTGGGCG  GGTATGGGGT  GCTTTTGGCC  TGC7GCCTTC
51  GCGCTTTTTC  GCGGCAATGG  TCGT7TGGCC  TTTGTGGGCG  GTGGCGGGGT
101  ATGACGGTTT  GCGCTGGGCG  GCGGTGCTGT  CGGATGCCTA  TATGCTCAA
151  CCGTTGGCGT  GACCGGTATT  TCAGCGAGCG  GCAACCTGTG  TCGTGGTGCT
201  GCGTTTGGCG  GTGCGTGTGG  CGTGGGTGCT  GCGCGGGCTG  CGCTTTCCGG
251  GCGCGGCTTT  GGTGCTGGCG  CTGCTGATGC  TGCTTTTGT  GATGCCACG
301  TTGGTGGCGG  GCGTGGGCGT  GGTGGTCTGT  TTCGGGCGCG  ACGGCGTGTN
351  GTGCGGCGCG  TGGCAGGATA  CGCGGTATCT  GTTGTGTGAT  GGCAATGTGT
401  TTTTTCACCT  TCGCTGTGTT  GTCAGGCGCG  CATATCAGGG  GTTTGTGCAA
451  GTGCGTGGCG  ACGGCTGTCA  GACGCGACAG  ACATGGGCG  CGGGCGGCTG
501  CGGGCGGTTT  TGGGACATTG  AATGCGCGGT  TTTGCGCGCG  TGGCTGGCG
551  CGGGCGTGTG  CCGTGTCTTC  CTGTATTGTT  TTTGCGGGTT  CGGGCTGGCA
601  TTGCTGTGCG  GCGGCGACCG  TTAAGCCACG  GTGGAAGTGG  AAATTACCA
651  GTTGTGTCATG  TTGCAACTCG  ATATGGCGGT  TGCTTCGGTG  CTNCTGTGGC
701  TGGTGTNNGG  GGTAAACNGG  GCGGCGAGGT  TGCTGTATGC  GTGGTTCCGG
751  AGGGCGGCGG  TTTCCGATAA  GCGGCTTTCC  CCGTGTATGC  CGTGGCGCGC
801  GCAATCGGCT  GGGGAATATG  TGCTNCTGGC  GTTGGCGGCG  CGGGTGTGTT
851  CTGCTGTGCT  CCGTGTTCNT  TTGTTGGCAA  TTGTTGTGAA  AGCGTGGTGC
901  GCGGCGGAAT  CCGTGGCGGT  GTTATGGAA  AGTGAACGT  GCGAGCGGTT
951  GTGGAATACT  NTGGGCTTCT  CGGGCGGCGG  GGTGTATGCG  CGGGCGGTTT
1001  TGGGTTGGTT  GTATGCGGGG  GCGGCGGCGG  GGTGCGGGTG  GATGCGGGGG
1051  CTGATCTTTT  GATGCTGTAT  GGTGTCGCGG  GTTGTGCTTT  CGGGCGGGGT
1101  GTGCTGCTTT  NATCGCGAGT  GGAAGGCTTC  TTTGCGCGTG  CTGCGGCA
1151  TGTATGCGCT  GCTGGCGTAT  CCGTGTGCG  CAAAGATGTT  TTTATCAGCG
1201  TGNATGCAC  TGCGCGCGGA  TTAACGCGAG  GCGGCGGCGG  GTTGGGTGCG
1251  AAGCGGCTTT  CAGACGGCAT  GCGGCGATCA  GTTCCCGCTC  TTGAACCGCG
1301  CGTGGCGGCG  CGGTGTGACT  TTGGCGGCGG  CAACCTGCGT  GGGCGAATT
1351  GCGGCAACCT  TGTTCNTGTC  GCGTCCGAG  TGGCAGACGC  TGACGACTTT

```


1401 GATTATGCC TATNTGGGAC GCGCGGGTGA NGATAATTAC GCGCGGGCGA
1451 TGGTGCTGAC ATTGCTGTTG GCGGCCTTCG CGCTGGGTAT NTTCTGCTG
1501 TTGGACGGCG GCGAAGGCGG AAAACGGACG GAAACGTTAT AA

This encodes a protein having amino acid sequence <SEO ID 578>:

5	1	MDGRWVWQA	AFALLPSPFLG	VPAMVAPLMA	VAAAYDGLAWR	AVLSDAVYMLK
	51	LRGWRVQVQA	ATCVVLVPLG	VPAMVAPLAL	VRVAVDGLWR	LLMLPVPMT
	101	VVAGVGVYAL	FGADGLXWRG	WQTPYLLLYL	GNFFVLLVL	VRAAYQGFVQ
	151	VPAARLQTLG	TIGADGWRFF	WQTEPMLPR	WLAGQGVCLV	LYCFSGFGLA
	201	LLGGSRYAT	VEVEIJQLVM	FELDMVAVSA	LVLWLVQVGA	AAGLLVAVGFL
10	251	RAAVSDKVA	PMPPSPQSV	GEDTGVHSLA	AVXSVCCLFV	LAIIVKAMS
	301	AGESWRVIME	STQVQAVNT	XRSFASALTE	NAUGVGVYTA	AAARSASVLA
	351	LMFLFVFLV	AGAGVCLVL	QWZATSLI	LVAVGVYVLA	EPYAKVLSA
	401	KDALPFOYR	AAAGVANGV	QTKVQVLE	LKPLARVLEL	LAAYGVYLA
	451	ALFLPXSKE	WQTLTPLYA	YXSGRDXNY	ARAMVLTLLL	AAAPALGXFL
15	501	LDGSGGGKRT	ETL*			

ORF139a and ORF139-1 show 96.5% homology over a 514aa overlap:

20	orf139a.pep	MDGRRWAVWGAFALLPSAFILAMVVPAPLWAAVAYDGLAWRAVLSDAYMLKRLAMTVFQAA
	orf139-1	MDGRRWVWGAFALLPSAFILAMVVPAPLWAAVAYDGLAWRAVLSDAYMLKRLAMTVFQAA
	orf139a.pep	ATCVILVLPLGVPAWVILARLAFPGRAVLRLRLMLPFVMPETLVAGVGVLALFGADGLXWRG
	orf139-1	ATCVILVLPLGVPAWVILARLAFPGRAVLRLRLMLPFVMPETLVAGVGVLALFGADGLXWRG
25	orf139a.pep	WQDTPYLLLYGNVFEXFLVPLVRAAYQGFGVQPAARLQTATLTSAGAAWRRFWDIEMPLVLR
	orf139-1	RQDTPYLLLYGNVFENLPLVLRRAAYQGFGVQPAARLQTARTLTSAGAAWRRFWDIEMPLVLR
30	orf139a.pep	WLAGGVCVLVFLYCFSGFGLALLGGSRYATVEVEIYQLVMFELDMAVASVLVWLVLGVTA
	orf139-1	WLAGGVCVLVFLYCFSGFGLALLGGSRYATVEVEIYQLVMFELDMAVASVLVWLVLGVTA
35	orf139a.pep	AAGLLYAWFGRRAVSDKAVSPVMPSPPGSYGEYVLLFAAAAVXSVCCLFXLLAIIVKAWS
	orf139-1	AAGLLYAWFGRRAVSDKAVSPVMPSPPGSYGEYVLLFAAAAVLSVCCLFPFLAIIVKAWS
40	orf139a.pep	AGESWRVLMSETWQAVWNTRFLSAAAVYAAAVLGVVYAAAAARRSAMWRGLMLFPFMVSP
	orf139-1	AGESWRVLMSETWQAVWNTRFLSAAAVYAAAVLGVVYAAAAARRSAMWRGLMLFPFMVSP
45	orf139a.pep	VCYSAGVLLXPQTASLPLLLAMYALLAYPFVAKVDLSAXDALPEPYGRAAAGLGANGF
	orf139-1	VCYSAGVLLXPQTASLPLLLAMYALLAYPFVAKVDLSAWDALPEPYGRAAAGLGANGF
	orf139a.pep	QTACRITFLPKLARLRGLTLAAATCVGEFAATLFSKRKEWQTTLTIYAYIGRAGXDNY
	orf139-1	QTACRITFLPKLARLRGLTLAAATCVGEFAATLFSRPEWQTTLTIYAYIGRAGEDNY
50	orf139a.pep	ARANVLTLLAAAFALXFLLLDGGEGGKRTETLX
	orf139-1	ARANVLTLLAAAFALGIFLLLDGGEGGKOTETLX

Homology with a predicted ORF from *N. gonorrhoeae*

ORF139 shows 95.2% identity over a 189aa overlap with a predicted ORF (ORF139ng) from

55 *N.gonorrhoeae*:

	orf139.pep	AWSAGESRVLMSETHAVVNTLRFSAAA	30
	orf139ng	QSVGEYVLLFASVAVLSVCLFLPLSATVVKWAGESRRLVMESETQAVVNTLRFSAAA	327
60	orf139.pep	VYAAVGLGVVYAAPARRSAWMRLGMXPFMSPVCVSAGVLLLYPGWTASLPLLIAMYAL	90
	orf139ng	VFAAAVLGVVYAAARRLVMMRGLVFLPFMSPVCVSAGVLLLYPGWTASLPLLIAMYAL	387

orf139.pep	LAYPFVAKDVLASWDALPPDYGRAAAGLGANGFQTACRITFPFLKPA LRRLGLTAAATCV	150
orf139ng	LAYPFVAKDVLASWDALPPDYGRAAAGLGANGFQTACRITFPFLKPA LRRLGLTAAATCV	447
5 orf139.pep	GEFAATFLSRPEWQTLTLLIYAYLGRAGEDNYARAMVL	189
orf139ng	GEFAATFLSRPEWQTLTLLIYAYLGRAGEDNYARAMVLTLLSAFVCFILLDNDEGG	507

The complete length ORF139ng nucleotide sequence <SEQ ID 579> is predicted to encode a protein having amino acid sequence <SEQ ID 580>:

10	1	MDGRCAVVRG	AFSLPLSAFL	AVMVVAPLWA	VVAYDGLAWR	AVLSDAYMLK
	51	RLAWTVFOAA	ATCVLVLPPLG	VPVAVWLARL	AFPGRAVLVR	LLMLFFVMFT
	101	LVAGVGVIAL	FGADGLLWRG	RQDTPYLLLY	GNVFNLPVL	VRAAYQGFQF
	151	VPAARLQATAR	TLGAGAWRRF	WDIEMPVLRP	WLAGGVCLVF	LYCFSGFGLA
	201	LLLGGSRYAT	VEVEIYQVLM	FELDMAGASA	LVMLVLGVTA	AAGLLYAWFG
15	251	RRAVSDKAVS	PVMPSPPOVS	GEYVLLAFSV	AVLSVCCFLF	LSAIVVKWAS
	301	AGESRRVIME	SETWQAVWNT	LRFSAAAVFA	AAVLGVVYVA	AARLIVMMRG
	351	LVLFPMVSP	VCVSAGVLLL	YPGWNTASLPL	LLMAYALLAY	PFVAKDVLSA
	401	WDALPPDYGR	AAAGLGANGF	QTACRITFPL	LKPA LRRLGLT	LAATCVGEF
	451	AATLFLSRPE	WQTLTLLIYA	YLGRAGEDNY	ARAMVLTLLL	SAFVCFILL
20	501	LDNGEGGKRT	ETL*			

Further work revealed a variant gonococcal DNA sequence <SEQ ID 581>:

	1	ATGGATGGAC	GGTGTGGGCG	GGTACGGGGT	GCTTTTTCCT	TGCTGCCTTC
	51	GGCTTTTTTG	GGGTAAATGG	TCGTTGGCGC	TTTGTGGGCG	GTGGCGGGCG
	101	ATGACGGTTT	GGCGTGGCGC	CGGTCGCTGT	CGGATGCCCTA	TATGCTCAAA
25	151	GCTTTGGCGT	GACCGGTGTT	TCAGCGCGCG	GCACCTGCTG	TGCTGCTGCT
	201	GGCTTTGGCG	GTGCGCTGCG	CGTGGGTGCT	GGCGCGGCTG	GGCTTCCCGG
	251	GGCGGGCTTT	GGTGTGCGCG	CTGCTGATGC	TGCGCTTTGT	GATGCCCCAG
	301	CTGGTGGCGG	GGTGGGGCGT	GCTGGCTCTG	TTCGGGGCGG	ACGGGCTGTT
	351	GTGGCGCGCG	CGCGAGGATA	CGCCGTATCT	GTTGTGTGAC	GGCAATGTGT
30	401	TTTTCAACCT	GCCCGTGTG	GTCAGGGCGG	CGTATCAGGG	GGTTGCTCAA
	451	GTGCTGCGCG	CACGGCTTCA	GACGGCACGG	ACGTTGGGCG	CGGGGGCGTG
	501	CGCGGGGTTT	TGGGACATTG	AAATGGCCGT	TTTGGCGCGG	TGGCTTGGCG
	551	CGCGCGTGTG	CCTGTCTCTC	CTGTATTGTT	TTTGGGGGTT	CGGGCTGGCA
	601	TGCTGTGTGG	CGCGCAGCGG	TTATGCCAAG	GTCGAAGTGG	AAATTTACCA
35	651	GTTGGTTATG	TTGCAACTCG	ATATGGCGGG	GGCTTCGGCG	CTGGTGTGGC
	701	TGCTGTGGGG	GGTAAAGCGG	CGCGCAGGGT	TGCTGTATGC	GTGGTTCCGG
	751	AGCGCGCGCG	TTTCCGATAA	GGCGGTTTCC	CGCGTGTATG	CTCGCGCGCG
	801	GCAATCGGTG	GGGGAATATG	TATTTGCTGC	ATTTTGGGTG	GGCGGTGTGT
	851	CGGTGTGCTG	CCTGTTTCTC	TTGTGGGCAA	TGTTGTGAA	AGCGTGGTGC
40	901	CGCGCGGAAT	CGCGCGGTGT	GTTAATGGAA	AGTGAARCTG	GGCAGCCAGT
	951	GTGGAATACT	TLGCGCTTTT	CGCGCGCGCG	GGTGTTCGG	CGCGCGGTTT
	1001	TGGGTGTGTT	GTATCGCGCG	CGCGCGCGCG	GGCTGTGTGT	GATGCGCGGA
	1051	CTGGTGTGTT	TACCGTTTAT	GGTGTGCGCG	GTTTGTGTT	CGCGCGGGGT
	1101	GCTGCTGCTT	TATCCGGGGT	GGACGGCTTC	GTTACCGGTG	CTGCTGGGCA
45	1151	TGTATGCGCT	GCTGGCGTAT	CGGTTGTGTT	CARAAGATGT	TTTATCGGCC
	1201	TGGGATGCAC	TGCGCGCGGA	TTACGGCAGG	CGCGCGCGAG	GTTTGGGCGC
	1251	AAACGGCTTT	CAGACGGCAT	CGCGTATCAC	GTTCCCGCTC	TTGAAACCGG
	1301	CGTTGGCGCG	CGGTCTGACT	TGCGCGCGCG	CGACGTGTGT	GGGCGAATTT
50	1351	CGCGCAACCT	TGTTCTGTGC	CGGTCCGGAA	TGGCAGACGT	TGACGACCTT
	1401	GATTTATGCC	TATCGCGGCG	GTGCGGTGTA	GGACAAATAT	CGCGCGGCAA
	1451	TGCTGTTCAC	ATGCGGCTGT	TGCGGATATG	CGGTGTGCTAT	TTTCTGCTGT
	1501	TTGGACAACG	CGGAAGCCGG	aaaACGGACG	GAAAGCTTAT	AA

This corresponds to the amino acid sequence <SEQ ID 582; ORF139ng-1>:

55	1	MDGRCAVVRG	AFSLPLSAFL	AVMVVAPLWA	VVAYDGLAWR	AVLSDAYMLK
	51	RLAWTVFOAA	ATCVLVLPPLG	VPVAVWLARL	AFPGRAVLVR	LLMLFFVMFT
	101	LVAGVGVIAL	FGADGLLWRG	RQDTPYLLLY	GNVFNLPVL	VRAAYQGFQF
	151	VPAARLQATAR	TLGAGAWRRF	WDIEMPVLRP	WLAGGVCLVF	LYCFSGFGLA
	201	LLLGGSRYAT	VEVEIYQVLM	FELDMAGASA	LVMLVLGVTA	AAGLLYAWFG
	251	RRAVSDKAVS	PVMPSPPOVS	GEYVLLAFSV	AVLSVCCFLF	LSAIVVKWAS
60	301	AGESRRVIME	SETWQAVWNT	LRFSAAAVFA	AAVLGVVYVA	AARLIVMMRG
	351	LVLFPMVSP	VCVSAGVLLL	YPGWNTASLPL	LLMAYALLAY	PFVAKDVLSA
	401	WDALPPDYGR	AAAGLGANGF	QTACRITFPL	LKPA LRRLGLT	LAATCVGEF
	451	AATLFLSRPE	WQTLTLLIYA	YLGRAGEDNY	ARAMVLTLLL	SAFVCFILL
	501	LDNGEGGKRT	ETL*			

ORF139ng-1 and ORF139-1 show 95.9% identity over 513aa overlap:

	orf139ng	MDCRCWAVRGAFSLPSAFLAVMVVAPLWAAVYDGLAWRAVLSDAYMLKRLAWTVFQAA
	orf139-1	MDGRRVWVGAFALLPSAFLAVMVVAPLWAAVYDGLAWRAVLSDAYMLKRLAWTVFQAA
5	orf139ng	ATCVLVLPGLGVFVAMVRLARLAFEGRALVLRLLMLPFVMPPTLVAGVGVLAIFGADGLLWRG
	orf139-1	ATCVLVLPGLGVFVAMVRLARLAFEGRALVLRLLMLPFVMPPTLVAGVGVLAIFGADGLLWRG
10	orf139ng	RQDTFYLLLYGNVFFNLPLVLRVRAAYQGFQVPAARLQARTLGGAGAWRRFWDIEMPVLRP
	orf139-1	RQDTFYLLLYGNVFFNLPLVLRVRAAYQGFQVPAARLQARTLGGAGAWRRFWDIEMPVLRP
15	orf139ng	WLACCVCVLFLYICFSGFGLALLGGSRVATVEVEIYQLVMFELMDAGASALVWLVLGVTA
	orf139-1	WLAGGVCVLFLYICFSGFGLALLGGSRVATVEVEIYQLVMFELMDAVASVWLVLGVTA
	orf139ng	AAGLLYAWFGRRAVSDKAVSPVMPSPQSVGEYVLLAFSVAVLVSVCCFLPFLSAIVVKAWS
20	orf139-1	AAGLLYAWFGRRAVSDKAVSPVMPSPQSVGEYVLLAFSVAVLVSVCCFLPFLSAIVVKAWS
	orf139ng	AGESRRVLMESSETWQAVWNTRLRFSAAVFAAAVLGVVYAAAARLVWVRGLVFLPFMVSP
25	orf139	AGESRRVLMESSETWQAVWNTRLRFSAAVFAAAVLGVVYAAAARLSAWVRGLVFLPFMVSP
	orf139ng	VCVSAGVLLLYPGWTASLFLLLAMYALLAYFFVAKDVLSAWDALPPEYGRAAAGLGANGF
	orf139-1	VCVSAGVLLLYPGWTASLFLLLAMYALLAYFFVAKDVLSAWDALPPEYGRAAAGLGANGF
30	orf139ng	QTACRITFFLLKPALRRGLTAAATCVGEFARTLFLSRPEWQTLTLLIYAYLGRAGEDNY
	orf139-1	QTACRITFFLLKPALRRGLTAAATCVGEFARTLFLSRPEWQTLTLLIYAYLGRAGEDNY
35	orf139ng	ARAMVLTLLLSAFAVCIPLLLDNGEGGKQTETL
	orf139-1	ARAMVLTLLLSAFALGIFLLDDGGEGGKQTETL

Based on the presence of a predicted binding-protein-dependent transport systems inner membrane component signature (underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or

40 diagnostics, or for raising antibodies.

Example 70

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 583>:

	1	ATGGACGGCT	GGACACAGAC	GCTGTCGGCG	CAAAACCTGT	TGGGCATTTC
	51	GGCGCGCGCA	ATCATCTCTCA	TTCTGATTTT	AATCGTCAGA	TTCCGCATCC
45	101	ACGCGCTGCT	GACACTGGTC	ATCGTCAGCC	TGCTCAGCGG	TTTGGCAACC
	151	GGTTTGCCCA	CAGGCAGCAT	TGTCAAAGAC	ATACTGTGCA	AAAACCTCGG
	201	CGGCACGCT	GGCGCGCTGG	CGCTTCTGGT	CGGCTGGGG	GCGATGCTCG
	251	AACGTTTGGT	C...			

This corresponds to the amino acid sequence <SEQ ID 584; ORF140>:

50	1	MDGWTQTLSA	QTLGISAAA	IILILILVR	FRHALLTLV	IVSLLTALAT
	51	GLPTGSIVKD	ILVKNFGGTL	GGVALLVGLG	AMLERLV..	

Further work revealed the complete nucleotide sequence <SEQ ID 585>:

	1	ATGGACGGCT	GGACACAGAC	GCTGTCGGCG	CAAAACCTGT	TGGGCATTTC
	51	GGCGCGCGCA	ATCATCTCTCA	TTCTGATTTT	AATCGTCAAA	TTCCGCATCC
55	101	ACGCGCTGCT	GACACTGGTC	ATCGTCAGCC	TGCTCAGCGG	TTTGGCAACC
	151	GGTTTGCCCA	CAGGCAGCAT	TGTCACAGAC	ATACTGTGCA	AAAACCTCGG

201 CGGCACGCTC GGGCGCGTGG CGCTTCTGGT CGGCGTGGGC GCGATGCTCG
 251 GACGTTTGGT CGAAACATCC GGGCGGGCAC AGTGCCTGGC GGACGCGCTG
 301 ATCCGGATGT TCGGCGAAAA ACGGCGCACG TTCGCGCTGG CGGTTGCCCTC
 351 GCTGATTTTC GGCTTCCCGA TTTTCTTCGA TGGCGGACTA ATCGTCATCG
 401 TGGCCATCGT GTTCGCCACC GCAAGGGGCA TGAACAGGA CGTACTGCC
 451 TTCGCGCTTG CCTCCATCGG CGCATTTTCC GTACATGACG TCTTCTGCC
 501 GCGCCATCCG GGGCGGATTG CGGCTTCCGA ATTTACGGC GCGAACATCG
 551 GCCAAGTTTT GATTTTGGGT CTGCGGACCG CCTTCACTAC ATGGTATTTT
 601 AGCGGCTATA TGCTCGGCAA AGTGTGGGG CGACCATCC AGTGTCCCGT
 651 TCGCAACTG CTCACCGCGG CGACGCGAGA CACAGGCTTC CCGAAGAAC
 701 CTGCCAAGC AGAAGCGGTC GTCCGCATCA TGCTGATTC CATGCTGCTG
 751 ATTTTCTCGA ATACCGGCGT ATCGGCGCTC ATCAGCGAAA AACTCGTAAG
 801 TGGCGACGAA ACCTGGGCTC AGACGGCAAA AATACTCGT TCGACACGGA
 851 TGGCGCTTCT GATTTCGGTA TTGGTGGCAC TGTTTGTCTT GGGACGCAA
 901 CGCGGCGAAA CGCGGACGCG GTTGAAAAAA ACCGTGGACG CGGCATCGC
 951 CCGCGTCTGT TCGGTGATTC TGATTACCGG CGCGGGCGGT ATGTTCCGGG
 1001 GCGTTTTGCG CGCTTCGCGC ATCGGCAAGG CACTCGCGGA CAGCATGGCG
 1051 GATTTGGGCA TTCCCGTCCCT TTTGGCTGT TTCTTGTGCG CCTTGGCACT
 1101 GCGTATCGCG CAAGGTTGCG CAACCGTCCG CCGTGCACCC GCGCGCGCGC
 1151 TGATGGCTCC TGGCGTTGCC GCGCGCGGCT TTACCGACTG GCAGCTCGCC
 1201 TGATCGGTAT TGCGACGCGC GGCAGGTTCT GTGCGTTGCA GCCACTTCAA
 1251 CGACTCGCGG TCTCGGCTGG TGGCGGCT CTGCGACATC GCGTACCGA
 1301 CGACGCTGAA AACCTGCGAG GTCAACCAAA CCTCATCGC ACTCATCGC
 1351 TTTGCCTTGT CGGCACTGCT GTTCGCCATC GTCTGA

25 This corresponds to the amino acid sequence <SEQ ID 586; ORF140-1>:

1 MDGWTQLSA OTLLGISAAA IILILILIVK FRIHALLTV IVSLLTALAT
 51 GLPTGGSIVND ILVKNFGGTL GGVALLVGLG AMLGRLVETS GGAQSLADAL
 101 IRMFGEKRAP FALGVASLIF GFPIFFDAGL IVMLPIVFAT ARMKMDVLP
 151 FALASIGAFS VMHVFLEPHF GPFAASEFYG ANIGQVLILG LPTAFITWYF
 301 SGVYLGKVLG RTIIVVPPEL LSGGTQDNDL PKEPKAGTV VAIMLPMML
 251 IFLNTGSVAL ISEKLVSADE TWVOTAKIG STPIALLISV LVALFVLGRK
 301 RGSBSGALEK TVDGALAPVC SVILITGAGG MFGGVLRAS IGKALADSMA
 351 DLGIPVLLGC FLVALALRIA QGSATVALLT AAALMAPAVA AAGFTDWQLA
 401 CIVLATARGV GCSHFNDG FNLVGRLLDM DVPTLTKTWT VNQTLIALIG
 451 FALSALLFAL V*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF140 shows 95.4% identity over a 87aa overlap with an ORF (ORF140a) from strain A of *N.*

meningitidis:

40 orf140.pep 10 20 30 40 50 60
 MDGWTQLSAQTLLGISAAAIIILILILIVRFRIHALLTVIVSLLTALATGLPTGSIKVD
 orf140a MDGWTQLSAQTLLGISAAAIIILILILIVKFRHALLTVIVSLLTALATGLPTGSIKVD
 45 10 20 30 40 50 60
 orf140.pep 70 80
 ILVKNFGGTLGGVALLVGLGAMLERLV
 orf140a ILVKNFGGTLGGVALLVGLGAMLERLV
 50 70 80 90 100 110 120
 orf140a VLVKNFGGTLGGVALLVGLGAMLERLVETSGGAQSLADALIRMFGEKRAPFALGVASLIF

The complete length ORF140a nucleotide sequence <SEQ ID 587> is:

1 ATGGACGGCT GGACACAGAC GCTGTCCGCG CAACCCCTGT TGGGCATTTT
 51 GGGCGGGGCA ATCAATCTCA TCTGATTTT AATCGTCAAA TTCGCGATCC
 101 ACGGCGTGCT GACACTGGTC ATGCTCAGCC TGCTGACGGC TTTGGCAACC
 151 GGTTTGGCCA CAGCGAGCAT TGTCAACGAC GTACTGGTCA AAAACTTCGG
 201 CGGACGCTC GCGCGCGTGG CGCTTCTGGT CGGCGTGGC GCGATGCTCG
 251 GACGTTTGGT CGAAACATCC GGGCGGGCAC AGTGCCTGGC GGACGCGCTG
 301 ATCCGGATGT TCGGCGAAAA ACGGCGCACG TTCGCGCTGG GGTTCGCTTC
 351 GCTGATTTTC GGCTTCCCGA TTTTCTTCGA TGGCGGACTA ATCGTCATCG
 401 TGGCCATCGT GTTCGCCACC GCAAGGGGCA TGAACAGGA CGTACTGCC
 451 TTCGCGCTTG CCTCCATCGG CGCATTTTCC GTACATGACG TCTTCTGCC

-335-

501 GCCCATCCG GCCCGATTG CCGCTCCGA ATTTTACGGC GCGAACATCG
 551 GCCAAGTTTT GATTTTGGGT CTGCGACCG CTTCTATCAC ATGGTATTTTC
 601 AGCGGCTATA TGCTCGGCAA AGTGTGGGG CGCACCATCC ATGTTCCCGT
 651 TCCCGAAGCTG CTCAGCGCGG GCACGCAAGA CAACGACCTG CGGAAGAAGAC
 701 CTGCCAAAGC AGGAACGGTC GTCCGCATCA TGCTGATTCC CATGCTGCTG
 751 ATTTTCTCTGA ATACCGGGCT ATCGGCCCTC ATCAGCGAAA AACTCGTAAG
 801 TCGCGACGAA ACCTGGGTTT AGACGGCAA AATAATCGT TGACACCGA
 851 TCGCGCTCTT GATTTCCGTA TTGGTGGAC TGTGTTCTT GCGACCGAAA
 901 CCGCGGAAA GCGCGAGCG GTTGGAAGA ACCGCGAGG GCGCATCTGC
 951 CCGCGCTCTT TCGGTGATTG TGATTACCGT CCGCGGCGGT AGTTCGCGG
 1001 CGGTTTTCGG CGGTTCCGGC ATCGCGAAGG CACTCGCGA CAGCATGGCG
 1051 GATTTGGGCA TTCCCGTCTT TTGGGCTGT TTCTTGTGC COTGGCACT
 1101 CGGTATCGCG CRAAGTTTCG CAACCGTGC COTGACACC GCGCGCGCG
 1151 TGATGGCTCC TGCCGTGCG GCCGCGGCT TTACCGACT GCGAGCTGCG
 1201 TGATATCGTAT TGCGAAGCGG GCGAGGTTCT GTGCGTTGCA GCGACTCAA
 1251 CGACTCGCG TTCTGGCTGG TCGGCGCGCT CTTGGCATGC AAGCTACCGA
 1301 CCACGCTGAA AACCTGGACG GTCAACAAA CCCTCATCGC ACTCATCGGC
 1351 TTGCGCTGT CGCACTGCT GTTCGCATC GTCTGA

This encodes a protein having amino acid sequence <SEQ ID 588>:

20 1 MDGWTQTLSA QTLGISA AAA IILILILIVK FRIHALTLIV IVSLTALAT
 51 GLPTGSIVND VLVNFGSTL GGVALLVGLG AMLGRIVETS GGAQSLADAL
 101 IRMFGEKRAP FALGVASLIF GFELFDAGL TVMLPIVFAT ARMKQDYLE
 151 FALASIGAFS VMHVFLPFP GPAAASEFYG ANIGQVLILG LPTAFITWYF
 201 SGYMLGKVLG RTIHVPPEL LSGGTQDNPL KPAPAKAGTV VAIMLPMML
 25 251 IFLNTGVSAL ISEKLVSADE TWQTAKIIG STPIALLISV LVALFVLGRK
 301 RGESSALEKT VDGLAPVCS VILITGAGMFGVLRASGIGKALADSMADLGI
 351 DLGIPVLLGC FLVALALRIA QGSATVALTT AAALMAPAVA AAGFTDWQLA
 401 CIVLATAAGS VGC SHFNDG FVLVGRLLDM DVPTTLKTWT VNQTLIALIG
 451 FALSALLFAI V*
 30 ORF140a and ORF140-1 show 99.8% identity over a 461aa overlap:
 orf140-1.pep MDGWTQTLSAQTLGISA AAA IILILILIVK FRIHALTLIV IVSLTALATGLPTGSIVND 60
 orf140a MDGWTQTLSAQTLGISA AAA IILILILIVK FRIHALTLIV IVSLTALATGLPTGSIVND 60
 35 orf140-1.pep IIVNFGGTLGGVALLVGLGAMLGRIVETS GGAQSLADALIRMFGEKRAPFALGVASLIF 120
 orf140a IIVNFGGTLGGVALLVGLGAMLGRIVETS GGAQSLADALIRMFGEKRAPFALGVASLIF 120
 40 orf140-1.pep GPFPFFDAGLIVMLPIVFATARRMKQDVLFPFALASIGAFSVMHVFLPFPHPGPAAASEFYG 180
 orf140a GPFPFFDAGLIVMLPIVFATARRMKQDVLFPFALASIGAFSVMHVFLPFPHPGPAAASEFYG 180
 orf140-1.pep ANIGQVLILGLPTAFITWYFSGYMLGKVLGRTHVVPPELLSGGTQDNPLKPAPAKAGTV 240
 45 orf140a ANIGQVLILGLPTAFITWYFSGYMLGKVLGRTHVVPPELLSGGTQDNPLKPAPAKAGTV 240
 orf140-1.pep VAIMLPMMLIPLANTGVSALISEKLVSADETWQTAKIIGSTPIALLISV LVALFVLGRK 300
 orf140a VAIMLPMMLIPLANTGVSALISEKLVSADETWQTAKIIGSTPIALLISV LVALFVLGRK 300
 50 orf140-1.pep RGESSALEKTVDGLAPVCSVILITGAGMFGVLRASGIGKALADSMADLGI PVLLGC 360
 orf140a RGESSALEKTVDGLAPVCSVILITGAGMFGVLRASGIGKALADSMADLGI PVLLGC 360
 55 orf140-1.pep FLVALALRIAQGSATVALTTAAALMAPAVARAAGFTDWQLACIVLATAAGSVGC SHFNDG 420
 orf140a FLVALALRIAQGSATVALTTAAALMAPAVARAAGFTDWQLACIVLATAAGSVGC SHFNDG 420
 60 orf140-1.pep FVLVGRLLDM DVPTTLKTWT VNQTLIALIGFALSALLFAI 461
 orf140a FVLVGRLLDM DVPTTLKTWT VNQTLIALIGFALSALLFAI 461

Homology with a predicted ORF from *N.gonorrhoeae*

ORF140 shows 92% identity over a 87aa overlap with a predicted ORF (ORF140ng) from

65 *N.gonorrhoeae*;

orf140.pep	MDGWTQLLSAQTLTGISAAAIILILILIVFRIRALLTLIVSVLLTALATGLPTGSIVKD	60
orf140.ng	MDGRQTQLLSAQTLTGISAAAIILILILIVFRIRALLTLIVSVLLTALATGLPTGSIVND	60
orf140.pep	ILVKNFGGTLGGVALLVGLGAMLERLV	87
orf140.ng	VLVKNFGGTLGGVALLVGLGAMLRIVETSGGAOSADALIRMEFEKRAPFPAGVASLIF	120

The complete length ORF140ng nucleotide sequence <SEQ ID 589> was predicted to encode a protein having amino acid sequence <SEO ID 590>:

10	1	MDGRPTLSA	OTILGISAAA	ITJLILTVLK	FMRALTVTV	TASLTALAT
	51	GLPTSGIVND	VLVKNFGTUL	GGVGLVGLS	AMRLVIVETS	GQAQSLADAL
	101	IRMFGRKRAP	PAHVASVILF	GPFFIFEDAGL	IVNMLVIVAT	ARMKQDQVLE
	151	FALSVAGSAG	PHVYV.PHEF	GTAFSEFSEF	AGMGLVILF	LPATVITVPL
	201	SGYVIGVND	GVVGVVGVV	LSGPTGVND	GVVGVVGVV	GVVGVVGVV
15	251	FLINTGVSYAL	ISERKLVASL	VTWQAKMIG	STPVALVLSV	LAALLVLGRM
	301	RGESGSTLSE	TVGDALAPAC	SVTLTGAGAG	MGVGVILRAG	IGKALADSKA
	351	DLSIFVLVGC	FLVALALRIA	QGSATVATVL	AAALMAPAVA	ARGQTDVTVA
	401	CVTLATRAS	VGSCHFNDSG	VLGVRLSGMS	DVPPFTKLTG	VNQTILAFIG
	451	FAVSAHARSA				

20 Further work revealed a variant gonococcal DNA sequence <SEO ID 591>:

	1	ATGGACGGCC	GACACACAGC	CGTGTGCGGG	CAACCTGTGT	TGGGCATTTC
	51	GGGGGGGGCA	ATGATCTCCA	TTCCTGCGGG	AATCGTCGAA	TTCGCATGCC
	101	GGCGCGCTGC	GACACTGTGC	ATCGCAGGCC	TGCTGCAGCG	TTTGGCAACC
25	151	GGTTTGTGCCA	GACGACGATC	CGTCAACGAC	GTATCTGGTCA	AAAACCTGGG
	201	CGCACGCGTC	GGGGGGGTGG	CGCTCTGTGT	CGGCTGTGGC	GCAATGCTCT
	251	GAGCTTTGTGT	AGAAACATCC	GGGGGCGGAC	AGTGGCTGCG	GGACGGCGTG
	301	ATCGAGATGT	TGCGGGAARA	ACGCGACACG	TTCGCTGTGG	GGGTTGCTCT
	351	CGTGATTTCG	TGCGTTCGCGA	TTTCTTGGG	TGCGGAGCTA	ATCGCTGACG
30	401	TGCGCATGCT	ATTGCGGACG	GACGCGCGCA	TGACACAGCA	GTATCTGGCC
	451	TGCGCGCTGT	CGCGCGTGGC	CGACTTTCGC	GTGATCGAGC	TCCTCTGGCG
	501	GGCCGACATG	CGCGCTGCGA	CGGCTGTGCG	GGAGCTGCTG	GGAGCTGCTG
	551	CGACGGTTTC	GATTTTGGGT	CGTGGGCGAG	CTTCATCATC	ATGGTATTTC
	601	AGCGGCTATA	TGCTGGGCGA	AGTGTGGGAG	CGGCGCTACT	ATGTTCCGCT
35	651	TGCGACGACT	CTCAGCGGGT	GCGCGACAGA	CAGCGACCGG	CGAAGAAAC
	701	CTCGCAACAG	AGGAAAGGTC	TGCGGCGTCA	TGCTGTATCC	CATCTGCTGT
	751	ATTTCTTGCA	ATCGAGGGGT	ATGACGCTCT	ATCAGCGAAA	AAGCTGTAAG
	801	TGCGGACGAA	ACTTGGGGTC	AGAAGCGAAT	TATGATAGCT	TGACACCTGT
	851	TGCGCCTCTC	GATTTCTCGTA	TGGCGGCAC	TGTGGTCTGT	GGGACGCAAA
40	901	CGCGCGGAAA	GGGCGACGAC	TTGTGGAGAC	ACCGTGGAGC	GGCGACTGAG
	951	CGCGCGCTGT	TGCTGTGATC	TGATTATGAG	CGGGGGCGGT	ATGTTGGGCG
	1001	GGGTTTGGG	CGGCTCGGCG	ATGCGGACGG	CAGTGGCGA	CAGCATGGCG
	1051	GATTTGGGCA	TGCGGCTGCT	TTTGGGCTGC	TTCCTGTGCT	CGTGGGACAT
	1101	GGTATCGGCG	CGAGGTGCGC	ACACCGTGCG	GTGACGCGA	CGCGCGGGCG
45	1151	TGATGATGCT	CGCGGCGGCG	CGCGGCGGCG	CGCGGCGGCT	CGCGGCGGCT
	1201	GTATCTGATC	TGCGCAACGC	CGACGATGCT	CGGCTTGGCA	GCACATCTGA
	1251	CAGCTCGGAC	TGTTGGGCTGT	TGGCGGCGCT	CTTGAGATGCT	GAGCTACGGA
	1301	CAACGCGTGA	AGCATGGAGC	GTCAACAAAC	CCCTCATGCG	ATTCATCGCG
	1351	TTTGCCTTGT	CGGCTCATGCT	GTTTGGCGAG	CTGTGTA	

This corresponds to the amino acid sequence <SEO ID 592: ORF140ng-1>:

```

50 1 MDGRTOFLSA OTLILGISAAA TLLILVLFI FRIRALLTV TASLTLALAT
51 GLPTSDIVN VMNVLNRGOTL GAGUALLVIG AMLRGLVETS GQASLADAL
101 IRMFGRERAP FAPGVASLIF GFPIEFADGL IYMLFIVAT ARMKQDVL
151 IRLASVSGFS FAPVLPPEH GFASFEFG IAKSDVETL LEPATFVTH
55 201 GYVNGVDFI FAPVLPPEH GFASFEFG IAKSDVETL LEPATFVTH
251 ILFNTGVSYAL ISEKLVASD SVYOTAKMIG IYVLLVSI IAAALVLRKK
301 RRGSSSTLSL TVDGLAPAPC SVTITAGAGL MGVGLVRASG IGLQADMSA
351 DLGFVFLVGC FLVALALRKA GGSATVATL AARMAPAPV AGQTFDQWLA
401 CIVLALRAS VCSHFNDSG VYVGLLEDM DVPFTLKPTV VQDTLQIA
451 FSAHLRKA VV

```

60 ORF140ng-1 and ORF140-1 show 96.3% identity over 461 aa overlap:

orf140ng-1.pep MDGRTQTLSAQTLGGISAAAIILILILIVKFRIRALLTLVIASLLTALATGLPTGSIVND

	orf140-1	MDGWTQTLSAQTTLLGISAAAAILILILIVKFRHALLTLVIVSLTLATALGLPTGSIVND
	orf140ng-1.pep	VLVKNFGGTLGGVALLVGLGAMLGRIVETSGGAQSLADALIRMFGEKRAPFAPGVASLIF
5	orf140-1	ILVKNFGGTLGGVALLVGLGAMLGRIVETSGGAQSLADALIRMFGEKRAPFAPGVASLIF
	orf140ng-1.pep	GFPIFFDAGLIVMLPIVFATARRMKQDVLFPALASVGAFPSVMHVFLPPHPGPIAASEFYG
10	orf140-1	GFPIFFDAGLIVMLPIVFATARRMKQDVLFPALASVGAFPSVMHVFLPPHPGPIAASEFYG
	orf140ng-1.pep	ANIGQVLLILGLPTAFITWYFSGYMLGKVLGRAIHVPVPELLSGGTQDSDPKPEKAKAGTV
	orf140-1	ANIGQVLLILGLPTAFITWYFSGYMLGKVLGRITIHVPVPELLSGGTQDNLKPEKAKAGTV
15	orf140ng-1.pep	VAVMLIPMLLIPIANTGVSAISEKLVSADETWVQTAKMIGSTPVALLISVLAALLVLGRK
	orf140-1	VAVMLIPMLLIPIANTGVSAISEKLVSADETWVQTAKIIGSTPIALLISVLAALLVLGRK
20	orf140ng-1.pep	RGESGSTLEKTVDGALAPACSVILITGAGGMFGVLRASGIGKALADSMDLGI PVLLGC
	orf140-1	RGESGSALEKTVDGALAPACSVILITGAGGMFGVLRASGIGKALADSMDLGI PVLLGC
	orf140ng-1.pep	FLVALALRIAQGSATVALTTAAALMAPAVAAAGPTDWQLACIVLATAAGSVGCSHFNDG
25	orf140-1	FLVALALRIAQGSATVALTTAAALMAPAVAAAGPTDWQLACIVLATAAGSVGCSHFNDG
	orf140ng-1.pep	FWLVGRLLDMDVPTTLKTWTVNQTLIAPIGFSALLFAIV
	orf140-1	FWLVGRLLDMDVPTTLKTWTVNQTLIALIGFSALLFAIV
30	Furthermore, ORF140ng-1 is homologous to an <i>E. coli</i> protein:	
	gi 882633 (U29579) ORF_o454 [Escherichia coli] >gi 1789097 (AE000358) o454;	
	This 454 aa ORF is 34% identical (9 gaps) to 444 residues of an approx. 456 aa	
	protein GMP BACIL SW: P46832 [Escherichia coli] Length = 454	
	Score = 210 bits (529), Expect = 1e-53	
35	Identities = 130/384 (33%), Positives = 194/384 (49%), Gaps = 19/384 (4%)	
	Query: 88	ETSGGAQSLADALIRMFGEKRAPFAPGVASLIFGPIFFDAGLIVMLPIVFATARRMKQD 147
		E SGGA+SLA+ R G+KR A +A+ G P+FFD G I++ PI++ A+ K
40	Subject: 80	EHSAGAESLANVFSRKLGDKRTIAALTAAFFLGLIPVFFDVGFIILAPIIYGFAKAKTS 139
	Query: 148	VLFPALASVGAFPSVMHVFLPPHPGPIAASEFYGANIGQVLLILGLPTAFITWYFSGYMLGK 207
		L F L G +HV +P+P+P+AA+ A+IG +I+G+ +I GY K
	Subject: 140	PLKFLGPVAGIMLTGVHVAVPPHPGPVAAAGLLHADIGWLTIGIAIS-IPGVVGYFAAK 198
45	Query: 208	VILGRAIHVPVPELL-----SGGTQDSDPKPEKAKAGTVVAVMLIPMLLIPIANTGV 257
		++ + + E+L G T+ SD P A V +++IP+ +I T
	Subject: 199	ITNKRQYAMSVEVLEQMLAPASEEGATKLSDKINPFGVA-LVTSILIVIPAIIMAGT-- 255
50	Query: 258	SALISEKLVSADETWVQTAKMIGSTPFXXXXXXXXXXXXXXXXXXGRKRGESGSTLEKTVDGALA 317
		+S L+ + T ++IGS +RG S + AL
	Subject: 256	---VSATILMPSPHLLGLTLQLIGSPMVALMIALVLAFVLLALARGWLSLQHTSDIMGSAIP 312
	Query: 318	PACSVILITGAGGMFGVLRASGIGKALADSMDLGI PVLLGCFLVALALRIAQGSXXXX 377
		A VIL+TGAGG+FG VL SG+GKALA+ + +P+L F+++IALR +QGS
55	Subject: 313	TAAVILVITGAGGVFGKVLVESGVGKALANMLQIDLP LPAAFIISIALRASQGS--AT 370
	Query: 378	XXXXXXXXXXXXXXXXXGPTDWQLACIVLATAAGSVGCSHFNDSGFVLVGRLLDMDVPTTLK 437
		G Q + LA G +G SH NDSGFV+V + L + V LK
60	Subject: 371	VAILTGGLLSEAVMGLNPICQVLVTLAACFGGLGASHINDSGFVITVTKYGLSVADGLK 430
	Query: 438	TWTVNQTLIAPIGFSALLFAIV 461
		TFTV T++ F GF ++ +R++
	Subject: 431	TWTVLTILGFTGLITWCWAVI 454

Based on this analysis, including the identification of the presence of a putative leader sequence

65 (double-underlined) and several putative transmembrane domains (single-underlined) in the

gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 71

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 593>:

```

5      1  ..GATTTCGGCA TATCGCCCGT GTATCTTTGG GTTGCCGCCG CGTTCACAA
51     TTTGCTGTGC CGTGGGCGTG CCGACTCATC CGATGTGCGA CGCTTTGCAG
101    GCGTATTTTT TGCGGTATC GGACTGACTT CCTGCGGCTT TGCGGTTTC
151    AACCTTTTGG GCAGACACCA CGGGCGCAC. GTGTCCTGA TTCTCATCGG
201    CTGTATCGGG CTGATTCCAG TTGCCCATTT CCTCAACCCC GCTGCCGCGG
10     251   CTTTTCGCGC CGCGGAGCTG GTGCTGCAGC GTTATCTTCT GGCTGCGCGG
301    CGCGTATGTC CGGCTCTTTT TCTGCTCGGT ACGGCTGGA CGCTGATGTC
351    GTTGCGACGA GGTATTCGG CAGCATTTGC CTTGATGCTG CCTTGCCCG
401    TACTGATGTT TTTCCGTCG ..

```

This corresponds to the amino acid sequence <SEQ ID 594; ORF141>:

```

15      1  ...DFGISFVYLW VAAAFKHLLS FWAADSYDVA RFAGVFFFAVI GLTSCGFAGF
51     NPLGRHHGRX VVLLILGIGI LIPVAFHLPN AAAFAAAGL VHGYSILARR
101    RVIAASFLLG TGWTMLSLAA AYPAAFALML FLFVLMFRFP ..

```

Further work revealed the complete nucleotide sequence <SEQ ID 595>:

```

20      1  ATGCTGACCT ATACCCCGCC CGATGCCCGC CGGCCCGCCA AAACCCACGA
51     AAGCCCGTGG CTGCTGCTGT TGATGGGCTT TGCGTGGTTG TGCGCCGCGG
101    TGTTCCTCCA CGATTGTTGG AATCTCGAOC AACCTGCGGT CTATACGCGC
151    GTGGAAGCAC TGGCAGGACG CCCACACCCC TTGGTTGCCC ATCTGTTCGG
201    TCAAAACGAT TCGGCGATAC CGCCCGTGTA TCTTTGGGTT GCCGCGCGGT
25     251   TCAAAACATT GCGTCTGCGC TGCGGCTGCG ACTCATACGA TGCCGCGACG
301    TTTCGAGCGG TATTTTTCGG CGTTATCGGA CTGACTTTCT GCGGCTTTGC
351    CGGTTTCAAC TTTTTCGGCA GACACCGATG GCGGCGATTC GTCTGATCTC
401    TCATCGGCTG TATCGGCTGC ATTCCAGTTC CCGACTTTCT CAACCCGCGT
451    GCGCGCGGCT TTGCGCGCGC CGGACTGGTG CTGCACGGTT ATTCTTTGGC
501    TCGCGCGGCG GTGATTTGCG CCTCTTTTCT GCTCGGTACG GGCTGGACGC
30     551   TGATGTGCTT GGCAGCAGCT TATCCGCGAG CATTTGCCCT GATGCTGCCC
601    TTGCCCCTAC TGAATGTTTT CGTCCGCTGG CAAAGCAGGC GTTTGATGTT
651    GACGCGAGTC GCCTCACCTG CCTTTGCCCT GCGGCTTATG ACCGTTTACC
701    CGCTGCTCTT GGCACAAAAC CAGCCCGGCG TGTTCGCGCA ATGGCTCGAC
751    TATCACGTTT TCGTACGTTT CGGCGCGGTC CGGCAOGTTC AGACGCGATT
35     801   CAGTTTGTCT TACTATCTGA AAAACCTGCT TTGTTTTCGA TTGCGCGGCG
851    TGCCGCTGGC GGTTTGGAAG GTTTGCAGCA CGCGGCTGTT TCGACGCGAC
901    TGCGGATTTT TGCGGCTGCT CTGGATGCTT GCGGTTTGGT TGCTGCTTTGC
951    CGTCAATCCG CAGCGTTTTC AGGATTAACCT GGTCTGGCTG CTCTCCGCGC
100   1001   TGCCGCTGTT CGGCGCGGCG CACTGAGACA GCGTGGAGCG GCGCGCGCG
105   1051   GCGTTTGTCA ACTGGTTCCG CATTATGGGG TTCGAGACTG TTGCGCTGTT
110   1101   CCTGTGGACG GGCTTTTTCG CCAATGAATTA CGGCTCGCCC GCCAAGCTTG
115   1151   CGGAACGCGC GCGCTATTTC AGCCCGTATT ATGTTCTCTGA TATCGATCC
120   1201   ATTCCGATGG GCGTTGCCGT ACTGTTTACA CCCTTGTGGC TGTGGGCGAT
125   1251   TACCCGGAATA AACATACGCG GCAGGACAGC GGTTACCAAC TGCGGCGGAC
45     1301   GCGTTACCTT GACCTGGGCT TTGCTGATGA CGCTGTTCTT GCCGTGGCTG
135   1351   GACGCGGCGA AAAGCCACGC GCGGCTGCTC CGGAGTATGG AGGCATGCTG
140   1401   TTCCCGGAAA TTGAACAGGG AGCTTTTACA CGGCATCGCA TGTATCGGCA
145   1451   TAGCGGCGCG GCACTTGACG ACGCGGATTG TTTGAGACGA GTAACGACAA
150   1501   TTGCGCGCAC CGGCTGGGCA TGTAACAATG CGTACGCGCA TCGTCTCTCT
50     1551   GCCCCAAATG GCGGATGCGC CGCAAGGCTG GCACAGCGTT TGCGAGGGTG
1601   1601   CGGCTCGCGC CAACAAAGAC AGTAAGTTGC CACTGATACG GAAATCGGG
1651   GAAATATAT AA

```

This corresponds to the amino acid sequence <SEQ ID 596; ORF141-1>:

```

55      1  MLTYTPDPAR FPAKTHEKFW LLLIMAFWL WGVFVSHDLW NFDEPAVYTA
51     VEALAGSPTE LVARLFQGTQ FCIPFVYLWV AAAPKHLSP WAADSYDAAR
101    FAGVFFFAVIG LSCGFAFGFN FLGRHHGRSV VLLILGIGL IPVAFHLPN
151    AAFAAAGLV LHGYSILARR VIAASFLLGT GWTMLSLAA YPAAFALMLP
201    LPVLMFRFW QSRRLMLTAV ASLAFALPIM TVYFLLAKT QFALFAQWLD

```


5
 251 YHVFGTGGV RHVQTAFSLF YYLKNLLWFA LPALPLAVMT VCTRLESTDO
 301 WGLGVVWML AVLVLLAVNP ORFDNLVWL LPFLALFGAA QLSLRLGAA
 351 AFVNWGIMA FGLFAVFLWT GFFAMNYGW AKLAERAAFY SPYYVDIDP
 401 IPMAVAVLET PLWLWATIRK NIRGRQAVTN WAAGVTLTWA LLMTLELPWL
 451 DAAKSHAPVV RSMESLSPE LKRELSGDIE CIGGGGLH TRIVTQYGT
 501 LHRVGDVQC RYRIVLLPQN ADAPOGWQTW WQARPRNKD SKFALIRKIG
 551 ENH*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

10 ORF141 shows 95.0% identity over a 140aa overlap with an ORF (ORF141a) from strain A of *N. meningitidis*:

```

                                10      20      30
orf141.pep                      DFGISPVYLVWAAAFKHLLSWPAADSYDVA
15 orf141a  WNPDEFAVYTAVEALGSPSTPLVAHLFGQIDFGIPFVYLVWAAAFKHLLSWPAADPYDAA
              40      50      60      70      80      90
orf141.pep  RFAGVFFAVIGLTSCGFAGNFLGRHGRXVVLILIGCIGLIPVAHILNPAAAAFAAAGL
20 orf141a  RFAGVFFAVVGLTSCGFAGNFLGRHGRSVVLILIGCIGLIPVTPLNPAAAAFAAAGL
              100     110     120     130     140     150
orf141.pep  VLHGYSLARRRVIAASFLLTGWTLMSLAAAYFAAFALMLFLPVLIMFFRP
25 orf141a  VLHGYSLARRRVIAASFLLTGWTLMSLAAAYFAAFALMLFLPVLIMFFRPQSRRLMLTA
              160     170     180     190     200     210
orf141a  VASLAFALPLMTVYVPLLLAKTQFALQWLDHVFGTGQVRHIQTAFSLFYLLKNLLWF
              220     230     240     250     260     270

```

The complete length ORF141a nucleotide sequence <SEQ ID 597> is:

```

1  ATGCTGACCT ATACCCCGCC CGATGCCCGC CGGCCCGCCA AAACCCACGA
51 AAAGCCGTGG CTGTTGCTGT TGATGGCGTT TGCTGGTTTG TGGCCCGGCG
35 101 TGTTTTCCCA CGATTGTGTG AATCTGCAGC AACCTGCCGT CTATACGGCC
151 GTCGAAGCAC TGGCAGGCAG CCCCACCCCT TTGGTTGCCC ATCTGTTCCG
201 TCAAATCGAT TTCGGCATAC CGCCCGTGTA TCTTTGGGTT GCCGCCGCGT
251 TCAAAACATT GCTGTGCGCC TGGGCTCGCG ACCGCTATGA TGGCCGACGC
40 301 TTTGCCGGCG TGTTTTTCGC CGTGTGCGGA CTGACTTCCT CGCGCTTTGC
351 CGGTTTCAAC TTTTGGGCGA GACACACGGG GCGCAGCGTC GTCCGTGATT
401 TCATCGGCGT TATCGGCGTG ATTCCGACGC TACACTTTCT CAACCCGCGT
451 GCCGCCGACCT TTGCGCGCGC CGGACTGGTG CTGCACGGTT ATTCTTTGCG
501 TCGCCGCGCG GTGATTCGCG CCTCTTTCT GCTCGTACG GTTGCGACGC
551 TGATGTCGTT GGCAGCAGCT TATCCGCGCG CATTGCGCTT GATGCTGCC
45 601 CTGCCCGCTG TGATGTTTTT CGTCCGCTGG CAAAGCAGCG GTTTGATGTT
651 GACGCCAGTC GCCTCGCTTG CTTTGCCTCT GCCGCTTATG ACGGTTTACC
701 CGCTGCTCTT GGCAAAACAG CAGCCCGCGC TGTTCGCGCA ATGGCTGAC
751 GATCAGGTTT TCGGTACGTT CGGCGGCGTG CGGCACATTG AGACGGCATT
801 CAGTTTGTGT TACTACTCTGA AAAACCTGCT TTGGTTTGCA TTGCTCGCG
50 851 TGC CGCTGCG GGT TTGGAAC GTTTGC CGCA CGCGCTGTT TTGCAACGAC
901 TGGGGGATT TGGCGCTGCT CTGGATGCTT GCGCTTTTGG TGCTGCTTGC
951 CGTCAATCCG CAGCGTTTTC AGGATAACCT CGTCTGGCTG CTTCGCGCG
1001 TTGCCCTGTT CGGCGGCGCG CAACTGGACA GCCTGAGACG CGGCGGCGCG
1051 GCGTTTGTC AACTGTTTGG CATTATGGCG TTGCGAGCTG TTGCGGTTGT
55 1101 CCGTGTGACG GCTTTTTCG CCAATGAATTA CGCTGGCCCC GCCAAGCTTG
1151 CCGAATTCG CCGTATATCT AGCCGATAT ATGTCCTGA TATCGATCC
1201 ATTCCGATGG CGGTTGCGGT ACTGTTTACA CCTTTGTGCG TGTGGGCGAT
1251 TACCCGCAAA AACATACCGC GCAGGCGAGC GGTTACCAAC TGGCGGCGAG
1301 GCGTTACCTT GACCTGGGCT TTGCTGATGA CGCTGTTTCT GCGCTGGCTG
60 1351 GACGCGGCGA AAAGCCACGC GCCCGTGTCT CGGAGTATGG AGGCATCGCT
1401 TTCCCGGGAA TTAAACACGG AGCTTTCAGA CGGCATCGAG TGTATCGACA
1451 TAGCGCGCGG CGACCTACAC ACGCGGATTG TTTGACGACA GTACGCGACA
1501 TTGCGCGACC CGCTCGGCGA TGTACAATGC CGCTACCGCA CTGCTCGCTT
1551 GCCCAACAC GCGGATGCGC CGCAGGCTG GCAGACGCTG TGGCAGGCTG

```

1601 CGCGCCCGCG CAACAAGAC AGTAAGTTCG CACTGATACG GAAAACCGGG
1651 GAAATATAT TAAACAAC AGATTGA

This encodes a protein having amino acid sequence <SEQ ID 598>:

5 1 MLTYTPPDAR PPAKTHEKFW LLLLMFAFWL WPGVFSHDLW NPDEPAVYTA
51 VEALAGSPTP LVAHLFGQID FGIPPVYLWV AAFAKHLSP WAADPYDAAR
101 FAGVFFAVVG LSCGFAGFN FLGRHHGRSV VLLIGCIGL IPTVHFLNPA
151 AAFAAAGLV LHGYSLARRR VIAASFLGT GWTILSLAA YPAFAFALMLP
201 LPVLMFFRFP QSRRLMLTAV ASLAFALFLM TVYPLLAKT QPALFAQWLD
251 DHVFGTGGV RHQTAFLS YLKNLWFA LPALFLAVT VCRTLFSTD
301 WGLGVVWML AVLLVAVNP QSFQNLWV LPPLALFGA QLSLRGAAG
351 AFVNWFGIMA FGLFAVFLWT GFAMNYGWP AKLAERAAV SPYVDPDIDP
401 IPMAVAVLTP PLWLWAIIRK NIRGRQAVTN WAAGVTLTWA LLMTLEFLPW
451 DAAKSHAPVV RSMEASLSP LKRELSGIE CIDIGGGDLH TRIVWTQYGT
501 LPHRVGDVQC RYRIVRLPQN ADAFQWQTV WQGARPRNKD SKFALIRKGT
15 551 ENILKTTD*

ORF141a and ORF141-1 show 98.2% identity in 553 aa overlap:

orf141a.pep MLTYTPPDARPPAKTHEKFWLLLMFAFWLWPGVFSHDLWNPDEPAVYTAVEALAGSPTP
|||
orf141-1 MLTYTPPDARPPAKTHEKFWLLLMFAFWLWPGVFSHDLWNPDEPAVYTAVEALAGSPTP
20 orf141a.pep LVAHLFGQIDFGIPPVYLWVAAFAKHLSPWAADPYDAARFAGVFFAVVGLTSCGFAGFN
|||
orf141-1 LVAHLFGQIDFGIPPVYLWVAAFAKHLSPWAADPYDAARFAGVFFAVVGLTSCGFAGFN
25 orf141a.pep FLGRHHGRSVVLLIGCIGLIPTVHFLNPAFAAAGLVHGYSLARRRVIAASFLGT
|||
orf141-1 FLGRHHGRSVVLLIGCIGLIPVAHFLNPAFAAAGLVHGYSLARRRVIAASFLGT
30 orf141a.pep GWTILSLAAAYPAFAFALMLPLPVLMFFRFPWQSRRLMTAVASLAFALPLMTVYPLLAKT
|||
orf141-1 GWTILSLAAAYPAFAFALMLPLPVLMFFRFPWQSRRLMTAVASLAFALPLMTVYPLLAKT
35 orf141a.pep QPALFAQWLDHVFVGTGGVVRHIQTAFSLFYLYKNLLWFAFALPLAVMTVCRTLFSTD
|||
orf141-1 QPALFAQWLDHVFVGTGGVVRHVQTAFSLFYLYKNLLWFAFALPLAVMTVCRTLFSTD
40 orf141a.pep WGLGVVWMLAVLLVAVNPQRFQDNLWVLLPPLALFGAQLDSLRRGAFAFVNWFGIMA
|||
orf141-1 WGLGVVWMLAVLLVAVNPQRFQDNLWVLLPPLALFGAQLDSLRRGAFAFVNWFGIMA
45 orf141a.pep FGLFAVFLWTGFFAMNYGWPAKLAERAAVSPYVDPDIDPIMAVAVLTPPLWLWAIIRK
|||
orf141-1 FGLFAVFLWTGFFAMNYGWPAKLAERAAVSPYVDPDIDPIMAVAVLTPPLWLWAIIRK
50 orf141a.pep NIRGRQAVTNWAAGVTLTWA LLMTLEFLPWLDAAKSHAPVVRSMASLSP LKRELSGIE
|||
orf141-1 NIRGRQAVTNWAAGVTLTWA LLMTLEFLPWLDAAKSHAPVVRSMASLSP LKRELSGIE
55 orf141a.pep CIDIGGGDLHTRIVWTQYGTLPHRVGDVQC RYRIVRLPQNADAPQWQTVWQGARPRNKD
|||
orf141-1 CIDIGGGDLHTRIVWTQYGTLPHRVGDVQC RYRIVRLPQNADAPQWQTVWQGARPRNKD
orf141a.pep SKFALIRKGTENI
|||
orf141-1 SKFALIRKGTENI

Homology with a predicted ORF from *N.gonorrhoeae*

ORF141 shows 95% identity over a 140aa overlap with a predicted ORF (ORF141ng) from *N.gonorrhoeae*:

60 orf141.pep DFGISPVYLWVAAFAKHLSPWAADSYDVA 30
|||
orf141ng WNPAPPAVYTAVALAGSPTPLVAHLFGQIDFGIPPVYLWVAAFAKHLSPWAADPYDAA 126

orf141.pep	RFAGVFFAVIGLTS CGFAGFNFLGRHHGRXVVLIIIGCIGLIPVAHFLNPAFAAAFAAAGL	90
orf141.ng	RFAGVFFAVIGLTS CGFAGFNFLGRHHGRSVLIIIGCIGLIPVAHFFNPAFAAAFAAAGL	186
5 orf141.pep	VLHGYSLARRRVIAASFLLTGTWTLMSLAAAYPAFAALMLPLPLVMFFRP	140
orf141.ng	VLHGYSLARRRVIAASFLLTGTWTLMSLAAAYPAFAALMLPLPLVMFFRPQSSRRIMLTA	246

An ORF141ng nucleotide sequence <SEQ ID 599> was predicted to encode a protein having amino acid sequence <SEQ ID 600>:

10	1	MPSEAVSARF	LCEYLLHLAI	RPFLLTMLLT	YTFPDPARPA	KTHEKFWLLL
	51	LMFAFWLWFG	VFESHLWNFA	EPAVITAVEA	LAGSPTFLVA	HLISQTDGFI
	101	PPVILWVAAA	FKHLLSPWAA	HPYDAARFAG	VFAVIGLTS	CGFAGFNFLG
	151	HHHGRSVLTI	HIGCIGLIPV	HPFTNPAFAA	FAAGLVLEGS	YSLARRRVIA
	201	ASFLLTGTWT	LMSLAAAYPA	AFALMLPLPV	LMFFRPWQSR	RIMLTAVASL
	251	AFALPMTVY	PILLAKTQPA	LFAQWLNVHV	FGTFGGVRHI	QRAFSLFHYL
	301	KNLLWFAPPG	LPLAVVTVC	TRLFTDWDGI	LGIVWMLAVL	VLLAFNPQRF
	351	QDNLWLLPP	LALFGAAQLD	SLRRGAAAFV	NWFGIMAFGL	FAVFLMTGFF
	401	AMNYGWPAKL	AERAAVFSFY	YVPDIDPIM	AVAVLFTPLW	LWATRKNIR
	451	GRQAVTNWAA	GVTLLTWALL	TLFLPWLDA	KSHAPVVRSM	EASFSPELKR
	501	ELSDGIEICG	IGGGDLHTRI	VMTQYGLPH	RVGDVRCYR	IVRLPQNADA
	551	PQGWQTVWQG	ARPRNKDSKF	ALIRKIGENI	LKTTD*	

Further work revealed the following gonococcal DNA sequence <SEQ ID 601>:

25	1	ATGCTGACCT	ATACCCCGCC	CGATGCCCGC	CCGCCCGCCA	AAACCCACGA
	51	AAAACCGTGG	CTGCTCTGTT	TGATGGCGTT	TGCGTGGCTG	TGGCCGCGAG
	101	TGTTTTCOCA	CGATTPTGTG	AACTCTGCCG	AACTCGCGGT	CTATACCGCC
	151	GTGCAAGCAG	TGGCAGGCAG	CCCCACCCCT	TGTGTTGCCG	ATCTGTTCGG
	201	TCAAACOGAT	TTCCGCATAC	CGCCCGTGTA	TCTTTGGGTT	GCCGCCGCAT
	251	TCAAACATTT	GCTGTGCGCG	TGGCGAGCGC	ACCCGATAGA	TGCCGCAACG
	301	TTTGACAGCG	TATTTTGTTC	CGTTATCGGA	CTGACTTCTT	CGCGCTTTGC
	351	CGGTTTCAAC	TTTTTGGGCA	GACACCAAGG	GCGCAGCGTT	GTTTAAATCC
	401	ATATCGGCTG	TATCGGCGTG	ATTCCGGTTG	CCCATTTCTC	CAATCCGcgcc
	451	gcgcgcgcct	tTGC CGCGCG	CGGACTGGTG	CTGCAcgctc	actcgctgGC
	501	ACGCGCGCGC	GTGATTgcgc	cctctTtccT	GCTCGGTACG	GSTTGGACGT
	551	TGATGTGCTC	GGCGCGAGCT	TATCCGCGCG	CGTTTGCCTC	GATGCTGCCC
	601	CTGCCCGTGC	TGATGTTTTT	CGTCCGCTGG	CAAAAGCAGC	GTTTATGTTT
	651	GACGCGAGTC	GCTCTGGCTG	CTTTTGCCCT	GCGCTTATFG	ACCGTTTACC
	701	CGCTGCTGCT	gcGCAAAACG	CACGCCGCGC	TGPTTGGCGA	ATGGCTCAAC
	751	TATCACGTTT	gcgtTACGTT	cgcgTGCCTG	CGGCAcATTG	AGAggGcAtT
	801	Cagtttggttt	cactatctgA	AAAatctgct	ttggttcgca	cgcgcgcggc
	851	TGCGCTGGCG	GGTTTGGACG	GTTTGC CGCA	CACGCTGTGT	TTCCAGCCAG
	901	TGGGGGATTT	TGGGCATGTG	CTGGATGCTT	GCGGTTTTGG	TGCTGCTCGC
	951	CTTTAATCCG	CAGCGTTTTT	AAGACAACCT	CGTCTGGCTG	CTGCCCGCGC
	1001	TTGCCCTGTT	CGCGCGCGCG	CAACTGGACA	GCTCGAGGCG	CGCGCGCGCG
	1051	GCTTTTGTC	ACTGGTTTCG	CATTATGGCG	TTCCGCGCTG	TTCCGCTGTT
	1101	CTCTGGGACG	GGCTTTTTCG	TCGATGAATTA	CGGCTGGCCC	GCCAAGCTTG
	1151	CGGAACGCGC	CGCTACTTTC	AGCCCGTATT	ACGTTTCCGA	CATCGATCCC
	1201	ATTCCGATGG	CGGTTGCGGT	ACTGTTTACA	CCCTTTGGGC	TGTGGCGGAT
	1251	TACCCGGAAA	ACATATACGG	CGAGCGCAGC	GTTTACCAAC	TGGCGGGCAG
	1301	GCGTTTACCTT	GACCTGGGCT	TGTGTGATGA	CGCTGTTTCT	GCGTGGGCTG
	1351	GACGCGCGCA	AAACGACGCA	CCCGCTGCTC	CGGAGTATGG	AGGCACTGTT
	1401	TTCCCGCGAA	TTAAACGCGA	AGCTTTCAGA	CGGCTACGCA	TGATATCGCA
	1451	TAGCGCGCGC	CGACCTTGAC	ACGCGATTGG	TTTGGACGCA	GTACGACGCA
	1501	TTGCCGCAAC	GCGTGGCGGA	TGTCGCTTGC	CGCTACCGTA	TGTCGCGCTC
	1551	GCCCCAAAAC	GCGGATGCGC	CGCAAGGCTG	CGACACGGTC	TGGCAGGGTG
	1601	CGCGCCCGCG	CAACAAGAC	AGTAAGTTTG	CAGTATGAC	GAATAFCGGG
	1651	GAATAATAT	TAAAAACAC	AGATTGA		

This corresponds to the amino acid sequence <SEQ ID 602; ORF141ng-1>:

60	1	MLTYTPPDAR	PPAKTHEKPW	LLLLMAFWL	WFGVFESHLW	NPAEPAVYTA
	51	VEALAGSPTP	LVAHLFGQTD	FGIPFPVILWV	AAAFKILLSP	WAADPYDAAR
	101	FAGVFFFAVIG	LTS CGFAGFN	FLGRHHGRSV	VLIIHIGCIGL	IPVAHFLNPA
	151	AAAFFAAGLV	LHGYSLARRR	VIAASFLLTG	GWTLMSLAAA	YPAFAALMLP
	201	LPVLMFFRPW	QSRRLMTAV	ASLAFALPLM	TVYPLLAWT	QFALFAQWLN
	251	YHVGTFGGV	RHIGRAFLSF	HYLKNLWLA	FGLFLAVMT	VCRRLFSTD
	301	WGILGIVWML	AVLVLLAFNP	QRFCQNLVWL	LPLALFGAA	QLDLSLRGAA

351	AFVNWFGIMA FGLFAVFLWT GFFAMNYGWP AKLAERAAYF SPYVVDIPDP
401	IPMAVAVLFT PLWLWATR K NIRGRQAVTN WAAGVTLTWA LLMTLFLPWL
451	DAAKSHAPVV RSEMEASFPE LKRELSDGIE CIGIGGGDLH TRIVWTQYGT
501	LPHRVGDVRC RYRIVRLPQN ADAPQGWQTV WQGARPRNKD SKFALIRKIG
551	ENLTKTID*
5	
ORF141ng-1 and ORF141-1 show 97.5% identity in 553 aa overlap:	
10	orf141ng-1.pep MLTYTPDPARPPAKTHEKFWLLMLMAFAWLWPGVFSHDLWNPAEPVAVTAVEALAGSFTT orf141-1 MLTYTPDPARPPAKTHEKFWLLMLMAFAWLWPGVFSHDLWNPAEPVAVTAVEALAGSFTT
	orf141ng-1.pep LV AHLFGQTD FGI PPVYLWVAAAFKHL LSPWAADPYDAARFAGVFFAVIGLTS CGFAGFN orf141-1 LV AHLFGQTD FGI PPVYLWVAAAFKHL LSPWAADSYDAARFAGVFFAVIGLTS CGFAGFN
15	orf141ng-1.pep FLGRHHGRSVVLIHIGCIGLIPVAHFLNFAAAAFAGLVLHGYS LARRRVIAASFLLGT orf141-1 FLGRHHGRSVVLIHIGCIGLIPVAHFLNFAAAAFAGLVLHGYS LARRRVIAASFLLGT
20	orf141ng-1.pep GWTIMSLAAAYPAAFALMLPLPVMFFRFWQSRRLMTAVASLAFALPLMTVYPLLAKT orf141-1 GWTIMSLAAAYPAAFALMLPLPVMFFRFWQSRRLMTAVASLAFALPLMTVYPLLAKT
25	orf141ng-1.pep QPALFAQWLNHYHVGTFGGVRHIIQRAFSLHYHLKILLWFAPPGPLPLAVWTVCRTRLFSTD orf141-1 QPALFAQWLDYHVGTFGGVRHIVQTAFSLFYLLKNLLWFALPALPLAVWTVCRTRLFSTD
30	orf141ng-1.pep WGILGIVWMLAVLVLLAFNPQRFDNLVWLLPPLALFGAQLDSLRGAAAFVNWFGIMA orf141-1 WGILGIVWMLAVLVLLAFNPQRFDNLVWLLPPLALFGAQLDSLRGAAAFVNWFGIMA
	orf141ng-1.pep FGLFAVFLWTGFFAMNYGWPAKLAERAAYFSPYVVDIPDIPMAVAVLFTPLWLWATR K orf141-1 FGLFAVFLWTGFFAMNYGWPAKLAERAAYFSPYVVDIPDIPMAVAVLFTPLWLWATR K
35	orf141ng-1.pep NIRGRQAVTNWAAGVTLTWWALLMTLFLPWLDAAKSHAPVVRSEMEASFPELKRELS DGIE orf141-1 NIRGRQAVTNWAAGVTLTWWALLMTLFLPWLDAAKSHAPVVRSEMEASFPELKRELS DGIE
40	orf141ng-1.pep CIGIGGGDLHTRIVWTQYGTLP HRVGDVRCRYRIVRLPQNADAPQGWQTVWQGARPRNKD orf141-1 CIGIGGGDLHTRIVWTQYGTLP HRVGDVRCRYRIVRLPQNADAPQGWQTVWQGARPRNKD
45	orf141ng-1.pep SKFALIRKIGENILKTTDX orf141-1 SKFALIRKIGENIX

Based on the presence of several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 72

50 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 603>:

```

1  ..CAATCCGCCA AATGGTTATC GGGCCAACT CTAGTCGGCA CAGCAATTGG
51 GATACGCGGG CAGATAAAGC TTGGCGGCAA CCTGCATTAC GATATATTTA
101 CCGGCCCGGC ATTGAAAAAG CCGCAATTTT TCCAATCAAG GAAATGGGCA
151 AGCGGTTTTT AGTAGGCTA TACGTTTTAA

```

55 This corresponds to the amino acid sequence <SEQ ID 604; ORF142>:

```

1  ..QSAWLSGQT LVGTAIGIRG QIKLGNNLHY DIFTGRALKK PEFFQSRKWA
51 SGFOVGYTF*

```

Further work revealed the complete nucleotide sequence <SEQ ID 605>:

-343-

	1	TTCCATCAAT	CGCGTAGTCA	CGCGCAGGGA	MAATATCCAG	GAATTAATCA
	5	TTTCTCTGCC	GACAGTCCT	TGGCGATGAT	TGATGTATTC	TATGTAAATC
	10	ATGCACTGTC	TGATGGCGCT	CACTGCGATG	AGCAAAAGTT	TACACGCCAT
	15	CGCAAGAGAC	CGGGATCAAA	CAATATGCCG	GTACATATT	CAGCCCTTTT
	20	CGGTAAATGC	ACATGGCTCAT	TCATACACAA	TGGCTACCTT	TACCATCAGG
	25	CAGTCTCCGC	ATTATCGGAA	TGATATGAT	ATGATGGAAA	AGTTTACAT
	30	ACTGATTCGT	CTCTCAACGC	CGTCTTGAT	CSTGATGCCA	AACGCCAAAC
	35	CTATCTCGCT	GTAATACTGT	GAGTGGAGCA	TACATTGATG	
	40	ATGCGCAACT	GAGTGTACAA	CGGGCTATAA	GTTTCGAGAT	
	45	CTTCTCCACA	AGACTATAT	CGTGCGCATG	ACGGCAGATT	TTAGTTTGAA
	50	ATATAAACGC	GACACCGGCA	TGAAGATGTC	TTTCGGCGCG	CTGGAAGAAG
	55	CTTCTGGGGA	AGGCACGTC	GTATGAAAA	TTTCGACGCG	ATCGCTGAT
	60	GTAATATCT	CTTTTCAAT	CGGTAAATCA	CTATTGCGCT	ATCAACATCT
	65	GTTCTATGCA	CAATGACACA	AAACCGCGCT	ACACATCGCA	AAATCAACAT
	70	CTATCGCGCG	ACGACAGAC	CTACCTGGCT	CTACGGGTA	ATATGATTGT
	75	TTCGCGGACG	CTGCGCGAC	CTGCGCGAC	GATTTTCTGA	GGCATATGGA
	80	ACCAACGCCAT	CAGCTTTATC	TTGGCGCGAT	TGTAGGACAT	GTTTCCAGAT
	85	ATACGCGCAA	ATGTTTATCG	GGCCAAACTC	TAGTGTCCAG	AGCAATTTCG
	90	ATAACGCGGC	AGATAAGGCT	TGGCGGCGAT	CTGCACTGAC	ATATATTTC
	95	CGCGCGCGCA	TGTGAAAAAG	CGGAATTTT	CCCATCAGG	AAATGGCCAA
	100	GCGGTTTCA	GGTAGGCTAT	CGATTTTAA		

This corresponds to the amino acid sequence <SEQ ID 606; ORF142-1>:

25	1	MNMSGSEATG	KYQGNITPFA	DNPLGLSDMF	VYVNGRSJGG	TPDEESFDCH
	5	KDEGGSNNTA	VOASAPFGKW	TWAFHNRGYS	YHQAVSLGE	VYDNGKSNY
	101	TRGFNRLLY	RDASRRTYLG	WKLWHTKTS	YTDUDELQY	RRKTGWLAE
	151	LSHKEYIGRS	TADFPLKYR	GTNKKDALRA	PEEFAEGES	RMKIWTASAD
	201	WNFPQIGTG	LPAFYDTSVA	QWMPPTLSQ	DLRAIGHHT	VRGFDGEMSL
	251	SAERGWWGN	DLSWQTFPGH	QLYGAVDG	VSGGSAKWL	GQTLVGTATC
	3	IRGOIKLGGN	LHYDITFGA	LKPEFPFSR	KWASGROW	TF*

30 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.gonorrhoeae*

ORF142 shows 88.1% identity over a 59aa overlap with a predicted ORF (ORF142ng) from *N.gonorrhoeae*:

35	orf142.pep	QSAKWLSSQTLVGTGATGIRGQIKLGGNLYH	30
	orf142.ng	RGWYWRNDLSWQFKPGHLYLGADVGHVSGQSAKWLSSQTLVGTGATGIRGQIKLGGNLYH	313
	orf142.pep	DIFTGRALKKPEFFQSRKWSAGFQVGYTF	59
40	orf142.ng	DIFTGRALKKPEYFOTKKWVTGFQVGYTF	342

The complete length ORF142ng nucleotide sequence <SEQ ID 607> is:

	1	WTGGATAAATT	CGGCTMTGA	TGGCAGACGGA	AAGATCCAG	GAAGTATCAC
	5	TTTCTTCCTG	GGCGATCCCT	GGGACATGAG	TGAATATGTC	TATGTAAATT
	101	ATCGAGCGTTC	AATTTGGCGG	ACGGCGCATG	AGGAAATATC	TGACGCCGCT
45	151	CCCAAGACGA	CGCGATCAAA	CAATACGCC	GTACATTATT	CAGCCCTTAT
	201	CGCTAAATGG	CAAGTGGCAT	TCAATTCAAA	TGGCTACGCT	TACCATCAAG
	251	CGGTTCCTGG	ATTATCGGAA	GTCTATGGAG	ATAATGGAAA	AAGTTACGAC
	301	ACTGATTTCTG	CTCTCAACCG	CTCTGTTGTAT	CGATGATGCA	AACGCGNAAC
50	351	CTATCTACCT	GTAATAACTGT	GACGACGGGA	CCGAAAGATG	TACATTTTATC
	401	ATGCGCGACT	CTGCTACGAA	CGGCTGTAAA	CCGACGGGTT	GTTCGCGAGAA
	451	CTTTCCCAACA	AGGATATATG	OAGGTGCTGA	ACCGCGAGAT	TTAAGTTTGA
	501	ATATTAATGCA	AGGACGCGGA	TGAAGAAATGC	CTTGGCGCGC	CTCGAAGAA
	551	CTTTTTCGCGA	GCGCAGCTCA	CGTATGAAAA	CTTGGACGCG	ATCGGCTGAT
55	601	GTAAATATCT	CTTTTCAAA	CGGTAACGCA	CTATTTGGCT	ATGACATCATC
	651	CTTATCTGCA	CAATGGAACA	AAAGCCCGCT	AMCATCGGCA	GACAAACTGG
	701	CTCTCGGCGC	ACACCAACACC	CTGCTTGGCT	TGCGAGCGAT	AATGAGTTTG
	751	CTATCGGACG	GGGATGGSTA	TGGGCGATG	GATTTGAGCT	GCGCAATTTAA
	801	ACACGAGCTCT	CAGCTTTATC	TGGGGCTGTA	GTGAGGACAT	GTTTCAGGAC
60	851	AATCGCGCAA	ATGGTTATCG	GCGCAAACTCT	TGGCGGAC	AGCAATTGGG
	901	ATACGCGGCG	AGTATAAAGT	TGGCGGACAC	CTTGTCTTAC	ATAATTATTC
	951	CGCGCTGCTA	CTGAAAGACG	CGGAATATTC	TCCAGACGAG	AAATGGGTAA

1001 CGGGGTTTCA GGTGGTTAT TCGTTTGA

This encodes a protein having amino acid sequence <SEQ ID 608>:

```

1 MDNSGSEATG KYQGNITFSA DNPFLSDMF YVNYGRSIGG TPDEENFDGH
51 RKEGGSSNNA VHYSAPFGKW TWAFNHNGYR YHQAVSGLSE VYDYNKSYN
101 TDFGNRLLY RDAKRRTYLS VKLWTRTKS YIDDAELTVQ RKRRTGWLAE
151 LSHKGYIGRS TADFKLYKH GTMKDALAR PEEAFEGGTS RMKIWTASAD
201 VNTPFQIGKQ LFAYDTSVHA QWNKTPLTSQ DKLAIGGHT VRGFDGMSL
251 PAERGWYWRN DLSWQFKPGH QLYLGADVGH VSGQSAKWS GQTLAGTAIG
301 IRGQIKLGGN LAYDFTGRA LKPEYFQTK KWTGPGVQGY SE*

```

The underlined sequence (aromatic-Xaa-aromatic amino acid motif) is usually found at the C-terminal end of outer membrane proteins.

ORF142ng and ORF142-1 show 95.6% identity over 342aa overlap:

```

orf142-1.pep MDNSGSEATGKYQGNITFSDNPLGLSDMFYVNYGRSIGGTPDEESFDGHRKEGGSSNNA
orf142ng-1 MDNSGSEATGKYQGNITFSDNPLGLSDMFYVNYGRSIGGTPDEENFDGHRKEGGSSNNA
orf142-1.pep VHYSAPFGKWTWAFNHNGYRYHQAVSGLSEVYDYNKSYNTDFGNRLLYRDAKRRTYLG
orf142ng-1 VHYSAPFGKWTWAFNHNGYRYHQAVSGLSEVYDYNKSYNTDFGNRLLYRDAKRRTYLS
orf142-1.pep VKLWMRETKSYIDDAELTVQRRKTAGWLAEASHKEYIGRSTADFKLYKRGTMKMDALRA
orf142ng-1 VKLWMRETKSYIDDAELTVQRRKTAGWLAEASHKEYIGRSTADFKLYKHGTGMKMDALRA
orf142-1.pep PEEAFEGGTSRMKIWTASADVNTPFQIGKQLFAYDTSVHAQWNKTPLTSQDKLAIGGHT
orf142ng-1 PEEAFEGGTSRMKIWTASADVNTPFQIGKQLFAYDTSVHAQWNKTPLTSQDKLAIGGHT
orf142-1.pep VRGFDGMSLPAERGWYWRNDLSWQFKPGHQLYLGAADVGHVSGQSAKWSGQTLAGTAIG
orf142ng-1 VRGFDGMSLPAERGWYWRNDLSWQFKPGHQLYLGAADVGHVSGQSAKWSGQTLAGTAIG
orf142-1.pep IRGQIKLGGNLYDIPTGRALKKPEFFQSRKWASGPGVQGYTF
orf142ng-1 IRGQIKLGGNLYDIPTGRALKKPEYFQTKKWTGPGVQGYSF

```

In addition, ORF142ng is homologous to the HecB protein of *E. chrysanthemi*:

```

g11772622 (L39897) HecB [Erwinia chrysanthemi] Length = 558
Score = 119 bits (235), Expect = 3e-26
Identities = 88/346 (25%), Positives = 151/346 (43%), Gaps = 22/346 (6%)
Query: 2 DNSGSEATGKYQGNITFSDNPLGLSDMFYVNYGRSIGGTPDEENFDGHRKEGGSSNNAV 61
      DNSG ++TG+ Q N + + DN FGL+D +++ G S + + D + G
Sbjct: 230 DNSGQKSTGEEQLNGSLALDNVFLADQWIFISAGHS---SRFATSHDAESLQAG----- 280
Query: 62 HYSAPFGKWTWAFNHNGYRYHQAVSGLSEVYDYNKSYNTDFGNRLLYRDAKRRTYLSV 121
      +S P+G W +N++ RY + G S F +R++RD KT ++
Sbjct: 281 -TSPYGYWNLGYNYSQSRYRNTFINRDPMHSTGDSDFHRFSLRVVRFRDGTMTAIG 339
Query: 122 KLVWTRTKSYIDDAELTVQRRKTAGWLAEASHKEYIGRSTADFKLYKHGTGMKMDALRA 181
      R +Y++ L RK + +H + A F Y G + +
Sbjct: 340 TFSQRTGNLYLNGSLPSSSRKLSSVSLGVNHSQKLWGLATFNPTYNRGRVNLGSETDT 399
Query: 182 EEAFFEGGTSRMKIWTASADVNTPFQIGKQLFAYDTSVHAQWNKTPLTSQDKLAIGGHTV 241
      +++ E + WT SA P Y S++ Q++ L ++L +GG ++
Sbjct: 400 DKSADFPRAEFNKWTLASASYHFV---TDSITYLGSLYQYSARALYGEQLTLGGESSI 456
Query: 242 RGFDEGMSLPAERGWYWRNDLSWQFKP----GHQLYLGA--DVGHVSGQSAKWSGQTLAG 296
      RGF E RG YWRN+L+WQ G+ ++ A D GH+ +L+G
Sbjct: 457 RGF-REQYTSGNRGAYWRNELNWQAWQLPVLGNVTFMAAVDGGHLYNHKQDNSTAAISLW 515
Query: 297 TAIGIRGQIKLGGNLYDIPTGRALKKPEYFQTKKWTGPGVQGYSF 342
      A+G+ + L + G + P + P + Q V G++VG SF

```

Sbjct: 516 GAVGMTVASRW---LSCQVTVGWPISYPAWLOPDTMVVGYRVGLSF 558

On the basis of this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

5 Example 73

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 609>:

1 ATGCGGACGA AATGGTCAGC AGTGAGAACG TGCTTACTTG GcGGGACACC
51 GCGGACMTGC ATACCGCTTT GAACCTGTGG TACGGTTTGC AAAAACCCGA
101 ATTCTCTAT GCGCATGAAA CGGCTCATCT AGAGCGGATC AATTTCWCGG
151 AAGCGCAATT GCCCTTCTGT ATCGGACATC TGTCGGGACG CGGTAAAGCGCG
201 TAATTGTAAT ATCGGGAACGG TCTGTAATAT CCGGACGCCA ATTTCCATCAT
251 TGAGGCGGCG GAAAGATTGG GGTTTGTGG CGGACGAATC GCACGATGG
301 AAAAGAAATC CCGCGTCTGT ATTAGBAGCA AC...

This corresponds to the amino acid sequence <SEO ID 610: ORF143>:

```

15      1  MRTKWSAVRS CTWADTADID TALNLLYRLQ KLEFLYGDEN GHSDGINLXD
      51  EQLPLIMEQL SGSGKALLVD RNLGLANAN FHHEAAEELG LLAAEVAQME
     101  KKYRLLIKNN

```

Further work revealed the complete nucleotide sequence <SEO ID 611>:

20	1	TGCGGATCAA	CATTCTTCACT	ACAGGAGAGAA	TTATATTGCC	GCTCTGACTCC
	51	ATGTCGGTGA	TACCTATGCGG	TATCCACAGGA	TGCCGCCACT	CGCGGTAAAA
	101	CTTGATGTGA	CAGCTGTGTG	AGAACAGAGTG	CGGACGAAAT	GGTGACGACTG
	151	GAGAGGTGCG	TTCCTTGGCG	GGACACATGAT	GACATCGATA	CGCTTTGTGAA
	201	CTGTGTTGAC	GTCTTGCAAA	ACTTCGAAAT	CTCTTATGCG	GATGAAAACGG
25	251	GCATTGACA	CGGATCAAT	TTGTGCGTGG	AGCAATTGCC	GTTCCTGATG
	301	GACAAATTGT	CGCGGACGCG	TGGAGCTGTA	TGTGTCGATC	GGAACGCTGT
	351	GTATCTTGCC	AAACGCAATT	TCCATCATGA	GGGGCGCGAA	GAGTTGGGGTG
	401	TTGTGGGCGG	AGAGATGCGA	CAGATGAAGA	AGAAATATCG	GGTGTGCTTT
	451	AAGACACACC	TGTATATCAA	CATATGAACAT	TGGGGGGTGT	GGCATCTGATT
30	501	CGCTCGAGCG	TAATATGACT	TTGATGCGAT	CGGAGGAGAT	TCGAAATGAT
	551	TATATGCGGG	ATTCGCGGAT	CGGCGAAGCG	GGGATATGAT	GGCATTTGTT
	601	ACTTGTGAAA	GAATTTGTTT	ATGCGGCTAT	AGGCAACGAG	GGCTATGTT
	651					GGTAA
	701					

This corresponds to the amino acid sequence <SEO ID 612; ORF143-1>:

35

1	MESTLSIQAN	LYPRLTTPAGA	FYAVSSDAPS	AGKTLHLHSL	KADADEMVSS
51	EKLLTWADTA	IDTIDLNTLIA	RQKLEFLYIG	DENGHS DGIN	LSDEQLFLIM
101	EQLSGSGKAL	LVDNRGLNYA	NANFHEAEAE	GLGLAAEVA	QMEKKYRLIA
151	KNNLYINNNA	WGCVDSQSGS	ELTFPFYLYIG	STKFILVIGG	IPDLGKEAFV
201	TLNRLYLRRY	SNRV*			

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

40 ORF143 shows 92.4% identity over a 105aa overlap with an ORF (ORF143a) from strain A of *N. meningitidis*:

45 orf143.pep M R T K W S A V R S C T W A D T I D T A L N L N L Y R L K L E F L
10 20 30
orf143a G A F Y A V S S D X P S A G K T L L H S L L K A D A D E M S S E K L L T W A T A D I D T A L N L N L Y R L K L E F L
20 30 40 50 60 70
50 orf143.pep Y G D E N G H S D G I N L N D Q L P L L M E Q L S G S K A L L V D R N G L Y I A N A N F H H E A A E L G L L A A E
40 50 60 70 80 90

-346-

orf143a	YGDENGHS	SDGINLS	DEQLPL	LLMEQL	SGSGKAL	LVDNRGL	YLANANF	HHEAAEEL	GLLAAE
	80	90	100	110	120	130			
5 orf143.pep		100	110						
		VAQMEKKYRL	LLIKNN						
orf143a		VAQMEKKYRL	XIKNNLY	NNNANGV	CDPSGQS	ELTFFFLY	IGSTFKFIL	VIGGIPDL	GKEA
	140	150	160	170	180	190			

The complete length ORF143a nucleotide sequence <SEQ ID 613> is:

10	1	ATGGAATCAA	CANTTCACT	ACAGCAAAAT	TTATATCNC	GCCTGACTCC
	51	TGCCGGTGA	CTTTATGCCG	TATCCAGCGA	TGNCCTCAGT	CGCGGTAAAA
	101	CTTTGTTGCA	CAGCTGTG	TAAGCGGATG	CGACGGAAT	GGTNAGCAAT
	151	GAGAACTGCG	TTACTTGGGC	GGANACCGCC	GACATCGATA	CGCTTTGAA
	201	CCTGTTGTAC	CGTTTGCRAA	AACTCGAATT	CCTCTATGGC	GATGAAAACG
15	251	GTCATTGAGA	CGGCATCAAT	TTGTCGGACG	AGCAATTGCC	GTTGCTGATG
	301	GAACAATTGT	CGGCGAGCGG	TAAGGCGTTA	TTGGTCGATC	GGAACGGTCT
	351	GTATCTTGCC	AACGCCAATT	TCCATCATGA	GGCGCGGAA	GAGTTGGGGT
	401	TGTTGGCGGC	AGAAGTCGCA	CAGATGGAAA	AGAAATACCG	GCTGNNATT
	451	AAGAACAACC	TGTATATCAA	CAATAACGCT	TGGGCGGTTT	CGATCCTTC
20	501	CGGTGAGAGC	GAATTGACAT	TTTTCCCAT	GTATATCGGT	TCAACCAAT
	551	TTATTTTGTG	TATCGGCGGC	ATTCCTGATT	TGGGCAAGA	GCAATTTGTT
	601	ACTTTGGTAA	GGATNTTATA	CCNCCNTTA	CAGCAACCG	GTGTAAACT
	651	TGGGAGAGAG	GANGGGTTAT	GCAGCAATTA	TTGA	

This encodes a protein having amino acid sequence <SEQ ID 614>:

25	1	MESTKSLQAN	LYXRLTPAGA	FYAVSSDXPS	AGKTLHSL	KADADEMVSS
	51	EKLTLTWAXTA	DIDTALNLLY	RLQKLEFLYG	DENGHS	SDGINLSDEQLPLLM
	101	EQLSGSGKAL	LVDNRGLYLA	NANFHEAAE	ELGLLAAEVA	QMEKKYRLXI
	151	KNNLYNNNA	WGVCDPSGQS	ELTFFFLYIG	STKFILVIGG	IPDLGKEAFV
	201	TLVRXLYXKL	QQPRVPLGRS	XGLCSNY*		

30 ORF143a and ORF143-1 show 97.1% identity in 207 aa overlap:

	orf143a.pep	MESTKSLQANLYXRLTPAGAFYAVSSDXPSAGKTLHSLKADADEMVSSSEKLTWTWAXTA
	orf143-1	MESTLQLQANLYPRLTTPAGAFYAVSSDAPSAAGKTLHSLKADADEMVSSSEKLTWTWADTA
35	orf143a.pep	DIDTALNLLYRLQKLEFLYGDENGHS
	orf143-1	DIDTALNLLYRLQKLEFLYGDENGHS
40	orf143a.pep	NANFHEAAEELGLLAAEVAQMEKKYRLXIKNNLYNNNANGVCDPSGQS
	orf143-1	NANFHEAAEELGLLAAEVAQMEKKYRLXIKNNLYNNNANGVCDPSGQS
	orf143a.pep	STKFILVIGGIPDLGKEAFVTLVRXLY
45	orf143-1	STKFILVIGGIPDLGKEAFVTLVRILY

Homology with a predicted ORF from *N.gonorrhoeae*

ORF143 shows 95.5% identity over a 110aa overlap with a predicted ORF (ORF143ng) from *N.gonorrhoeae*:

50	orf143.pep	MRTKWSAVRSCTWADTADIDTALNLLYRLQKLEFLYGDENGHS	SDGINLXDEQLPLMEQL	60
	orf143ng	MRTKWSAVRSCTWADTADIDTALNLLYRLQKLEFLYGDENGHS	SDGINLXDEQLPLMEQL	60
55	orf143.pep	SGSGKALLVDNRGLYLANANFHEAAEELGLLAAEVAQMEKKYRLIKNN		110
	orf143ng	SGSGKALLVDNRGLYLANANFHEAAEELGLLAAEVAQMEKKYRLIKNN		120

An ORF143ng nucleotide sequence <SEQ ID 615> was predicted to encode a protein having amino acid sequence <SEQ ID 616>:

1 MRTKWSAVRS CSRADTADID TALNLLYRLQ KLEFLYGDEN GHSDGINLSD
 51 EOLPLLMEO LSGSGKALLVD RNLGLYLANAN FHHESAEELG LLAAEVAQME
 101 KKYRLLIRNN LYINNNAWGV CDPGSGQSELT FFPLYIGSTK FILVIAGIPI
 151 LSKGGICYEG KDFIPPLQOP RVKLGTGGIM ROLLISILED LNNTSTDIIA
 201 SAVISTDGLF MATMLPSHLN SDRVGAISAT LLALGSRVQ ELACGELEOV
 251 NIKGSGYIL LSQAGKDAVL VLVAKETGRL GLILLDAKRA ARHIAEAT*

Further work revealed the following gonococcal DNA sequence <SEQ ID 617>:

1 ATGGAATCAA CACTTTCAC TACAAGCGAAT TTATATCCCT GCGTCACTCC
 51 TGGCGGTGCA TTTTATGCGG TATCCAGCGA TCGCCCGCAT GCGGCTAANA
 101 CTTTCTTGGC CAGCGCTGTT AAGCGGATG CGGACGAGT GGTCAACAGT
 151 GAGAAGCTGC TCGCGSGGGA CACCGCGGAC ATCATACCGC CTTTGAACCT
 201 GTTGTACCGT TTGCAAAAAC TCGAATTCCT CTATGGCGAT GAAACCGGTC
 251 ATTGACACGG CATCAATTG TCGGACGAGC AATTGCGGTT GCTGATGGAA
 301 CAATTGTCCG CGACGGGTAA GGCATTATTG GTCGATCGGA ACGGTCTGTA
 351 TCTTGCCAA CCAATTTCC ATCATGAGTC GCGGAAAGAG TTGGGTTGT
 401 TGCGGCGAGA AGTCGCACAG ATGGAAGAAGA AATACGCGCT GCTGATTAGG
 451 AACAACTCTG ATATCAACAA TAACGCTTGG GCGGTTTGGC ATCTTCCGG
 501 TCAGAGCGAA TTGACATTTT TCCCATTTGA TATCGGTTC ACAAATTTA
 551 TTTTGGTTAT CGCGCGGATT CCGCATTTGA GCAAGAGGCG ATTTGTACT
 601 TTGTAAGGA TTTTATACCG CGGTACAGC AACCGGCTG AA

This corresponds to the amino acid sequence <SEQ ID 618; ORF143ng-1>:

1 NESTLSLOAN LYPCLTGAGA FYAVSSDAPS AGKTLRLSL KADADEVVSS
 51 EKLLAADTAD IDPALNLLYR LKLEFLYGD ENHSDGINL SDEQLPLME
 101 QLSGSGKALL VDRNGLYLAN ANFHHESAE LGLAAEVAQ MEKKYRLLIR
 151 NNLYINNNAW GVCDPGSGQSE LTFFPLYIGS TKFILVIAGI PDLKAEFVT
 201 LVRILYRRYS NRV*

ORF143ng-1 and ORF143-1 show 95.8% identity in 214 aa overlap:

orf143ng-1.pep MESTLSLOANLYPCLTGAGAFYAVSSDAPSAGKTLRLSLKADADEVVSSSEKLLA-ADTA 59
 30 orf143-1 MESTLSLOANLYPRLTPAGAFYAVSSDAPSAGKTLRLSLKADADEVVSSSEKLLTWADTA 60
 orf143ng-1.pep DIDTALNLLYRLQKLEFLYGDENHSDGINLSDEQLPLMEQLSGSGKALLVDRNGLYLA 119
 35 orf143-1 DIDTALNLLYRLQKLEFLYGDENHSDGINLSDEQLPLMEQLSGSGKALLVDRNGLYLA 120
 orf143ng-1.pep NANFHHESAEELGLAAEVAQMEKKYRLLIRNNLYINNNAWGVCDPSGQSELTFFPLYIG 179
 40 orf143-1 NANFHHEAEEELGLAAEVAQMEKKYRLLIRNNLYINNNAWGVCDPSGQSELTFFPLYIG 180
 orf143ng-1.pep STKFILVIAGIPDLKAEFVTLVRILYRRYSNRV 213
 40 orf143-1 STKFILVIGGIPDLKAEFVTLVRILYRRYSNRV 214

Based on the presence of the putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 74

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 619>:

1 ATGACCTTTT TACAACGTTT GCAAGGTTTG GCAGACAATA AAATCTGTGC
 51 GTTTCATGTT TGTGCTGTCG GCGCGCTTGA TGAAGAACGC GTACCGCAGT
 101 CGGCGGCAAG CATGACGTTT ACGACGCTGC TGGCATCTGT CCGCGTCTGT
 151 ACGGTGATGG TGCGGCTGCG TCGGATTTTC CCGGTGTTGC ACGCGTGGTC
 201 GGATTCGTTT GTCTCTCTTC TCAACAAAC CATTTGCGCC CA. GCGCGGG
 251 ACATGGTGTT CGACTATATC AATGCGTTTC GCGAGCAGCG GAACCGGCTG
 301 ACGGCAATCG GCACGCTGAT GCTGGTCGTT ACCTCGCTCA TGCTGATTCG
 351 GACGATAGAC AATACGTTCA ACCGCAVCTG G₂CGGGTCAA WTYCCAGCGT
 401 CCGTGGATG..

-348-

This corresponds to the amino acid sequence <SEQ ID 620; ORF144>:

```

1  MTFLLRQLQGL ADNKICAFAM FVVRFFDEER VPQXAASMTF TTLLALVPLV
51  TVMVAVASIF PVFDRWSDSF VSFVNQTIYP XGADMVFDYI NAFREQANRL
101 TAIGSVMLVU TSLMLIRTID NTFNRIWRVX XQRPWM...

```

5 Further work revealed the complete nucleotide sequence <SEQ ID 621>:

```

1  ATGACCTTTT TACAACTTTT GCAAGGTTTG GCAGACAATA AAATCTGTGC
51  GTTTCGATGG TTGCTCGTCC GCCGCTTTGA TGAAGAACCG GTACCGCAGG
101 CGGCGGCAAG CATGACGTTT ACCAGCGCTGC TGCACTCGT CCCTCGCTGC
151 ACCGTGATGG TGGCGGTGCG TTGCACTTTT CCCTGTGTCG ACCGTGTGTC
201 GGATTCGTTC GTCTCCTTCG TCAACCAAAC CATTGTGCGC CAGGCGCGCG
251 ACATGGTGTT GCATCATATC AATGCGTTCG GCGAGCAGCG GAACCGGCTG
301 ACGGCAATCG CGAGCGTGAT GCTGGTCTGT ACCTCGCTGA TGCTGATTGG
351 GAGCATGACG AATACGTTCA ACCGCTATCG GCGGTCATAT TCCGACGTCG
401 CTGGAGGATG GCACTTTCTC GTCTATGGCG CTTTACTGAC GTTCCGGGCG
451 CTGCTTTTGG CGGTGGGCGT TTCTTTATG GTGCGCTGCG TACAGGATGC
501 CGGCTTTGCC TCAGGTGCGC CGCAGTGGTC GGGCGCGTTG CGAACCGGCG
551 CGAGCGTGAC CTTCTAGACG CTTTGTCTGT GGGGCGTGA CCGCTTCTGT
601 CCAACCGCTC TCGTCCCGCG GCGGCGAGCG TTGTGCGGGG CTTTGGCAAC
651 AGCGCTTTGT CTGGAACCG CGCGCTCCCT CTTCACTTGG TATATGGGCA
701 ATTTGACGCG CTACCGCTCG ATTTACGCGC GTTGTGCGCG CGTGCCGTTT
751 TTTCTGTTGT GGGTGAACCT GTTGTGACG CTGCTCTTGG CGGCGCGGCT
801 GCTGACTTCT TCACTCTCCT ACTGGCAGGG AGAAGCGTTC CGCAGGGGCT
851 TCGACTCGCG CGGACGTTTT GACGACGTGT TGAAATCTCG CGTGCTCTGT
901 GATGCGGCGC AAAAAGAGGG CAAGCGCTTG CTTGCTCAGG AGTTCAGACG
951 GCATATCAAT ATGGGCTACG ACCAGTGTGG CGAGCTTTTG GAAAGCTTGG
1001 CGGCGGACGG CACATCATAT TCGCGCTAGC AGGTTGGGAT GTTGAACGCG
1051 GGGGCGGATT CGATGAGGTT GAGCAAGCTC TCGATGCTC TCGTGTACCT
1101 TCCGTTGCCT GTGGAAAGGG ATCATGTGAA CCAAGCTGTC GATGCGGTAA
1151 TGACACCGTG TTTCAGACT TGAACATGAA CGCTGCGAGA GTTTGACGCT
1201 CAGGCGAAAA AACGCGAGTA G

```

This corresponds to the amino acid sequence <SEQ ID 622; ORF144-1>:

```

1  MTFLLRQLQGL ADNKICAFAM FVVRFFDEER VPQXAASMTF TTLLALVPLV
51  TVMVAVASIF PVFDRWSDSF VSFVNQTIYP QGADMVFDYI NAFREQANRL
101 TAIGSVMLVU TSLMLIRTID NTFNRIWRVN SQRPMMQFLV YVWALLTFGP
151 LSLGVGISFM VGSVDAAALA SGAPQWSGAL RTAATLTFT LLLWGLYRFV
201 PNRFPARQA FVGLATAFC LETARSLFTW YMGNFDGYRS IYGAFAAVFP
251 FLLMLNLLMT LVLGVALTSS SLSYWGGEAF RRGFDSRGRF DDVLKILLLL
301 DAAQKEGKAL PQQEFRRHIN MGYDELGELL EKLAHGYIY SRRGQWVLT
351 GADSELNEL FKLFVIRFLP VERDHVNVAV DAVMTFLCQT LNMTLAEFDA
401 QAKKQ*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF144 shows 96.3% identity over a 136aa overlap with an ORF (ORF144a) from strain A of *N. meningitidis*:

```

45  orf144.pep 10 20 30 40 50 60
      MTFLLRQLQGLADNKICAFAMFVVRFFDEERVVPQXAASMTFTTLLALVPLVTMVAVASIF
      orf144a  MTFLLRQLQGLADNKICAFAMFVVRFFDEERVVPQXAASMTFTTLLALVPLVTMVAVASIF
              10 20 30 40 50 60
100  orf144.pep 70 80 90 100 110 120
      PVFDRWSDSFVSFVNQTIYPXGADMVFDYINAFREQANRLTAIGSVMLVVTSLMLIRTID
      orf144a  PVFDRWSDSFVSFVNQTIYPXGADMVFDYINAFREQANRLTAIGSVMLVVTSLMLIRTID
55  orf144.pep 70 80 90 100 110 120
      NTFNRIWRVXXQRPWM
      orf144a  NTFNRIWRVNSQRPWMQFLVYVWALLTFGFLSLGVGISFXVGSVDAAALASGAPQWSGAL
              130
              NTFNRIWRVNSQRPWM
              NTFNRIWRVNSQRPWMQFLVYVWALLTFGFLSLGVGISFXVGSVDAAALASGAPQWSGAL
60

```

130 140 150 160 170 180

The complete length ORF144a nucleotide sequence <SEQ ID 623> is:

1 ATGACCTTTT TACAACTTTT GCAAGGTTTG GCAGACAATA AAATCTGTGC
 5 51 GTTTGCATGG TFCGTCGTCC GCGCGCTTTA TGAAGAAACG GTACCGCAGG
 101 CGGCGGCAAG CATGACGCTT ACGACACTGC TGGCACTGCT CCGCGTCTGC
 151 ACGCGTATGG TGGCGGTGCG TFCGATTTTC CCGGTGTTCG ACCGNTGGTC
 201 GGATTCGTTT GTCTCCTTGG TCAACCAAAAC CATTGTGCGC CAGGCGCGCG
 251 ACATGGTNTT CGACTATATC AATCGCTTCC GCGACGAGCG GAACCGCGGT
 301 ACGGAAATAT GTCGTGCTAT ACCTCGGAGA TCGTGTATGG TCGTGTATGG
 10 351 GACGATAGAC AATACGTTCA ACGCACTGCT GCGGCTCAAT TCCGACGCTC
 401 CGTGATGATG CAGATTTTCT GTCTATTGGG CTTTACTACG GTTCGCGCGG
 451 CTGCTCTTGG GCGTGGGCAT TTCCTTTATN GTCGCGTCGG TACAGGATGC
 501 GCGCTGTGCG TCAGGTGCGC GCGAGTGGTG GGGCGGCTGA CGACGCGCGG
 551 CGACGCTGAN CTTCATGACG CTTTGTCTGT GGGGCTGTA COGCTNCGTG
 15 601 CCAACCGCTC TCGTCCCGC GCGGCANGCG TTTGTGCGGG CTTTGGCAAC
 651 AGCGTCTGT CTGGAACCG GCGGTTCCCT CTTTACTTGG TATATGGGCA
 701 ATTCGACGCG CTACCGCTCG ATTTACGNGC CGTTTGCOCG CGTGCGCTTT
 751 TTTCTGTGTG GGCTGAACCT GTTGTGAGCG CTGCTCTTGG GCGGCGCGGT
 801 GGTGACTTCT TCACTCTCCT ACTGGCAGGG AGAAGCGGTC CAGCGGNGCT
 20 851 TCGACTCGCG CGGACGCTTT GACGACGCTG TGAATACTCT GCTGCTTCTG
 901 GATGCGCGCG AAAAAGAGGG CNAAGCCTTG CTGCTTCAGG AGTTTCAGAG
 951 GATATCATAT ATGGGCTACG ACGAGTTGGG CGAGCTTTTG GAAAGCTGCG
 1001 CGGCGACAGG CTACATCTAT TCCGCGACAG ACGGTTGGGT GTTGAACAG
 1051 GGGCGGATTT CGATTGAGTT GAACGAACCT TTCAGCTCTC TCGTTTACCG
 25 1101 TCGCTTGCTT GTGGAAAGGG ATCATGTGAA CCAAGCTGTC GATCGGGTAA
 1151 TGATGCCGTG TTTGACAGCT TTGAACATGA CGCTGCGAGA GTTTGACGCT
 1201 CAGGCGAAAA AACAGCAGCA ATCTTGA

This encodes a protein having amino acid sequence <SEQ ID 624>:

1 MTFLLRLQL ADNKICAFAW FVVRFRDEER VPQAAASMTF TFLALVLPVL
 51 TVMVAVASIF PVFDRWSDSF VSFVNQITVP QGADMVFDYI NAFREQANRL
 101 TAIGSVMLVV TSMXLIRITD NTFNRIWRVN SQRPMMQPL VYWALLTFGP
 151 LSLGVGISFX VGSVQDAALA SGAPQWSGAL RTAATLXFMT LLLWGLYRXV
 201 ENRFVPARKA FVGALATAPC LETARSLFTW YMGNFDGYSR IYGAFAAVFP
 251 FLILNLMT LVLGGAVLTS SLSYQGEAF RXPDSRGRF DDVLKILLLL
 301 DAAQKEGKAL PVQFRRRHIN MGVDLSEL LKARHGYIY SGRQGWVLT
 351 GADSTEINEL FKLFPVRLP VERDHNQAV DAVMMPCLQT LNTLAEFDA
 401 QAKKQQQS*

ORF144a and ORF144-1 show 97.8% identity in 406 aa overlap:

40 orf144a.pep MTFLLRLQLADNKICAFAWFVVRFRDEERVVPQAAASMTFTFLALVLPVLTVMVAVASIF
 orf144-1 MTFLLRLQLADNKICAFAWFVVRFRDEERVVPQAAASMTFTFLALVLPVLTVMVAVASIF
 45 orf144a.pep PVFDRWSDSFVSFVNQITVPQGADMVFDYINAFREQANRLTAIGSVMLVVTSMXLIRITD
 orf144-1 PVFDRWSDSFVSFVNQITVPQGADMVFDYINAFREQANRLTAIGSVMLVVTSMXLIRITD
 50 orf144a.pep NTFNRIWRVNSQRPMQPLVYWALLTFGPLSLGVGISFXVGSVQDAALASGAPQWSGAL
 orf144-1 NTFNRIWRVNSQRPMQPLVYWALLTFGPLSLGVGISFXVGSVQDAALASGAPQWSGAL
 55 orf144a.pep RTAATLXFMTLLWGLYRXVFNRFVPARKAFVGALATAPCLETARSLFTWYMGNFDGYSR
 orf144-1 RTAATLXFMTLLWGLYRXVFNRFVPARKAFVGALATAPCLETARSLFTWYMGNFDGYSR
 60 orf144a.pep IYGAFAAVPPFLLNLNLWTLVLGGAVLTSSLSYQGEAFRXPDSRGRFDDVLKILLLL
 orf144-1 IYGAFAAVPPFLLNLNLWTLVLGGAVLTSSLSYQGEAFRXPDSRGRFDDVLKILLLL
 65 orf144a.pep DAAQKEGKALPVQFRRRHINMGYDELGSLEKLAHGYIYSGRQGWVLTGADSTEINEL
 orf144-1 DAAQKEGKALPVQFRRRHINMGYDELGSLEKLAHGYIYSGRQGWVLTGADSTEINEL
 orf144a.pep FKLFPVRLPVERDHNQAVDAVMMPCLQTLNMTLAEFDAQAKKQQQS 408
 orf144-1 FKLFPVRLPVERDHNQAVDAVMMPCLQTLNMTLAEFDAQAKKQQ 406

Homology with a predicted ORF from *N.gonorrhoeae*

ORF144 shows 91.2% identity over a 136aa overlap with a predicted ORF (ORF144ng) from *N.gonorrhoeae*:

5	orf144.pep	MTFLQRLQGLADNKICAFAMFVVRDFEERVPOXAAASMTFTTLALVPLVTVMVAVASIF	60
	orf144ng	MTFLQWQGSADNKICAFAMFVIRRFSEERVPOXAAASMTFTTLALVPLVTVMVAVASIF	60
10	orf144.pep	PVFDWRSDSFVSFVNQITVPXGADMVFYIDAFRDQANRLTAIGSVMLVVTSLMLRTID	120
	orf144ng	PVFDWRSDSFVSFVNQITVPQGADMVFYIDAFRDQANRLTAIGSVMLVVTSLMLRTID	120
	orf144.pep	NTFNRIWRVXXQRPWM	136
15	orf144ng	NAPNRIWRVNTQRPWMQFLVYWALLTFGPLSLGVGISFMVGSVQDSVLSGGAQQWADAL	180

The complete length ORF144ng nucleotide sequence <SEQ ID 625> is predicted to encode a protein having amino acid sequence <SEQ ID 626>:

20	1	MTFLQWQGS	ADNKICAFAM	FVIRRFSEER	VPQAAASMTF	TTLALVPLV
	51	TVMVAVASIF	PVFDWRSDSF	VSFVNQITVP	QGADMVFYI	DAFRDQANRL
	101	TAIGSVMLV	TSMLMLRTID	NAPNRIWRVN	TQRPWMQFL	VYWALLTFGP
	151	LSLGVGISFM	VGSVQDSVLS	SGAQQWADAL	KTAARLAFMT	LLNLGLYRFV
	201	PNRFVPAQQA	FVGLALITAF	LETARFLFTW	YMGNFQGYRS	TYGFAAIVPF
	251	FLNLNLNLLWT	LVLGGAVLTS	SLSYWQGEAF	RRGFDSRGRF	DDVLKILLLL
	301	DAQKEGRTL	SVQEFRRHIN	MGYDELGELL	EKLARYGYIY	SGRGQWVLT
25	351	GAUSIELSEL	FKLFVYRPLP	VERDHVQNAV	DAVMTPCLOT	LMTLAEFDA
	401	QAKKQOOS*				

Further work revealed the following gonococcal DNA sequence <SEQ ID 627>:

30	1	ATGACCTTTT	TACACGCTTG	GCAAGGTTTG	GCGGACAATA	AAATCTGTGC
	51	ATTTCGATCG	TTGCGTACCC	GCGCTTTCAG	TGAAGACGCG	GTACCGCAGG
	101	CAGCGGGGAG	CATGACGTTT	ACGACACTCG	TGGCACTCGT	CCCGCTACTG
	151	ACCGTAATGG	TGCGGGTCCG	TTGATTTTC	CCCGTGTCCG	ACCGCTGGTC
	201	GGATTCTGTC	GTCCTCTTCG	TCAACCAAC	CATTGTGCGC	CAGGGCGCGG
	251	ATATGCGTGT	CGACTATATC	GACGATATCC	CGCATCAGCG	AAACCGGCTG
	301	ACCGCCATCG	GCAGCGTGAT	GCTGTGCTGA	ACCTCGCTGA	TGCTGATTCC
35	351	GACGATAGAC	AATGCGTTCA	ACCGCATCTG	GCGGGTTAAC	ACGCAACGCC
	401	CCTGGATGAT	GCAGTTCCCT	GTTTATTGGG	CGTTCGTCAG	TTTCGGGCCCT
	451	TTGCTCTTGG	GTGTGGGACT	TTCTTTATAG	GTCCGGTCGG	TTCAAGACTC
	501	CGTACTCTCC	TCCGGAGCGC	AACAATTGGC	GAGACGCTTG	AAGACGCGCG
	551	CAAGGCTGGC	TTTCATGACG	CTTTTGCTGT	GGGGCTGTGA	CCGCTCTGTG
40	601	CCCAACCGCT	TCGTGCCCGC	CCGGCAGCGC	TTTGTGCGAG	CTTTGATTAC
	651	GGCATTTCTG	CTGGAGAGCG	CAGCTTTCCT	GTTCACTCGG	TATATGGGCA
	701	ATTTCGACGG	CTACCGCTCG	ATTTCAGCGC	CATTTCGCGC	CGTCGCGTTT
	751	TTCTCGCTGT	GGTTAAACCT	GCTGTGAGCG	CTGGTCTTGG	GCGGGGCGGT
	801	GCTGACTCTG	TGCTGTGCTT	ATTGGCAGGG	CGAGGCGCTT	CGCAGGGGAT
45	851	TGCACTGCGC	CGGACGCTTT	GACGACGTGT	TGAATACTCT	GCTGCTTCTG
	901	GATGCGCGCG	AAAGACAGGG	CGAACCTCG	TCGCTTCAGG	AGTTGAGACG
	951	GCATATCAAT	ATGGGTTTACG	ATGAATTGGG	CGAGCTTTTG	GAAAGCTGGC
	1001	CGCGGTACGG	CTATATCTAT	TCCGCGCAGC	AGGGCTGGGT	TTTGAACACG
	1051	GGGCGGGATT	CGATTGAGTT	GAGCGAAGCT	TTCAAGCTCT	TCGTGTACCG
50	1101	CCCGTTGCGT	gtggaAAGGG	ATCATGTGAA	CCAAGCTGtc	gaTGCGGTAA
	1151	TGAGcccgctG	TTTGCAAGCT	TTGAACATGA	CGCTGGCGGA	GTTTGACGCT
	1201	CAGcgAAAA	AACAGCAGCA	GTCTTGA		

This encodes a variant of ORF144ng, having the amino acid sequence <SEQ ID 628; ORF144ng-1>:

55	1	MTFLQWQGL	ADNKICAFAM	FVIRRFSEER	VPQAAASMTF	TTLALVPLV
	51	TVMVAVASIF	PVFDWRSDSF	VSFVNQITVP	QGADMVFYI	DAFRDQANRL
	101	TAIGSVMLV	TSMLMLRTID	NAPNRIWRVN	TQRPWMQFL	VYWALLTFGP
	151	LSLGVGISFM	VGSVQDSVLS	SGAQQWADAL	KTAARLAFMT	LLNLGLYRFV
	201	PNRFVPAQQA	FVGLALITAF	LETARFLFTW	YMGNFQGYRS	TYGFAAIVPF
	251	FLNLNLNLLWT	LVLGGAVLTS	SLSYWQGEAF	RRGFDSRGRF	DDVLKILLLL
60	301	DAQKEGRTL	SVQEFRRHIN	MGYDELGELL	EKLARYGYIY	SGRGQWVLT

351 GADSIELSEL FKLFVYRFLP VERDHVQAV DAVMTFCLQT LNMTLAEFDA
401 QAKKQQQS*

ORF144ng-1 and ORF144-1 show 94.1% identity in 406 aa overlap:

```

5  orf144ng-1.pep  MTFLQRWQGLADNKICAFWVIRRFSEERVPPQAAASMTFTLLALVPVLTVMVAVASIF
    orf144-1      MTFLQRWQGLADNKICAFWVIRRFSEERVPPQAAASMTFTLLALVPVLTVMVAVASIF

10 orf144ng-1.pep  VFVDRWSDSFVSFVNQTIIVPQGDMMVFYDIADFRDQANRLTAIGSVMLVVTSLMLRTID
    orf144-1      VFVDRWSDSFVSFVNQTIIVPQGDMMVFYDIADFRDQANRLTAIGSVMLVVTSLMLRTID

15 orf144ng-1.pep  NAFNRIWRVNTQRPNWMMQFLVYWALLTFGLPLSLGVGISFMVGSVQDSVLSSGAQQWADAL
    orf144-1      NTFNRIWRVNSQRPNWMMQFLVYWALLTFGLPLSLGVGISFMVGSVQDAAALASGAPWWSGAL

20 orf144ng-1.pep  KTAARLAFMTLLWGLYRFVFNRFVPAQAFVGLALITAFCLTARFLFTWYMGNFDGYRS
    orf144-1      RTAATLTFTMTLLWGLYRFVFNRFVPAQAFVGLALITAFCLTARSLFTWYMGNFDGYRS

25 orf144ng-1.pep  DAAQKEGRTLSVQEFRRHINMGYDELGELLEKLARGYIYISGRQGVWLKTGADSIELSEL
    orf144-1      DAAQKEGKALPVQEFRRHINMGYDELGELLEKLARGHYIYISGRQGVWLKTGADSIELNEL

30 orf144ng-1.pep  FKLFVYRFLPVERDHVQAVDAVMTFCLQT LNMTLAEFDAQAKKQQQS
    orf144-1      FKLFVYRFLPVERDHVQAVDAVMTFCLQT LNMTLAEFDAQAKKQQ

```

On this basis of this analysis, including the identification of several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

35 Example 75

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 629>:

```

1  ..AGACACGCCCGCGCATCCG CATCGACACC GCCATCAACC CGAACTCGGA
51  AGCCCTCGCC GAACACCTCC ACTACCAATG CGAGGGCTTC CTCTGGCTCA
101 GCACCGATAT GCGTCAGGAA ATTTCGCCCC TGCTCATCCT GCTGCAACGC
40  151 ACCCGCGCGA AATGGCTGGA TGCCCAACGAA CGCCARACACC TGCGCCAAAG
201 CCGCTGTGAA ACACGGGAAC ACGGCTCGA

```

This corresponds to the amino acid sequence <SEQ ID 630; ORF146>:

```

1  ..RHARRIRIDT AINPELEALA EHLHYQWQGF LWLSTDMRQE ISALVILLQR
51  TRRKWLDAHE RQHLRQSLLE TREHG*

```

45 Further work revealed the complete nucleotide sequence <SEQ ID 631>:

```

1  ATGAACACCT CGCAACGCAA CCGCTCGTC AGCGCTGCG TCAACTCCTA
51  CGAACCGCTAC CGCTACCGCC GCGTCATCCA CGCGTCCGG CTCGGCGGGG
101 CGTCCCTGTT CGCCACCGCC TCGCCCGCG TGCTCCACCT CCACACCGGC
50  151 GAGTGGATAG GAGTACCGT CTTCGCTGTC CTGCGATGC TCAGATTCA
201 AGGGGCGATT TACTCCAAGG CGGTGGAAAG TATGCTCGGC ACGGTCATCG
251 GCGCTGGGCGC GGGTTTGGGC GTTTTATGCG TGAACACGCA TTATTTCCAC
301 GGCACCTCC TCCTTACCT CACCGTCGCG ACGGCAAGCG CACTGGCGCG
351 CTGGCGCGCG GTCCGCAAAA ACGGCTACGT CCTATGCTG GCAGGGCTGA
401 CGATGCTGTAT GCTCATCGGC GACAAACGCA CGCAATGGCT GCACAGCGGA
55  451 CTCATGCGCG CCATGAACGT CCTCATCGCG GCGGCCATCG CCATGCGCGC

```

501	CGCCAAACTG	CTGCCGCTGA	AATCCAACT	GATGTGGCGT	TTCAATGCTTG
551	CGGACAACCT	GGCCGACTGC	AGCAAAATGA	TTGCCGAAAT	CAGCAACGGC
601	AGGCGCATGA	CCCGCGAACG	CCTCGAGGAG	AACATGCGCA	AAATGCGCCA
651	AATCAAACGA	CGCATGGTCA	AAAGCCCGAG	CCATCTCGCC	GCCACATCGG
701	GGGAAAGCGG	CATCAGCCCC	GCCATGATGG	AAGCCATGCA	GCAAGCCAC
751	CGTAAATTCG	TCAACACCAC	CGAGCTGCTC	CTGACCAACG	CGGCCAAGCT
801	GCAATCTCCC	AAACTCAACG	GCAGGAAAT	CGGCTGCTT	GACGCCACT
851	TCACACTGCT	CAACAACGAC	CTGCACAAA	CGCTCGCCCT	TATCAACGGC
901	AGCACACCTC	GGCGATCCG	CATCGACAC	GCATCAAC	CGCACTCGA
951	AGCCCTCGCC	GAAACACCTC	ACTTCCAATG	GCAAGGCTTC	CTCTGGCTCA
1001	GCAACCAATAT	GGCTCAGGAA	ATTTCGCCCT	TGCTCATCTC	GCTGCAACGC
1051	ACCCGCGCA	AATGGCTGGA	TGCCCAAGAA	GCACAACCC	TGCGCCAAAG
1101	CCTGCTTGAA	ACACGGGAAC	ACGGCTGA		

This corresponds to the amino acid sequence <SEQ ID 632; ORF146-1>:

1	MNTSQRRRLV	SRWLSYERY	RYRRLHAVR	LGGAVLFATA	SARLLHIQHC
51	EWIGMTVFVV	LGMLQFGGAI	YSKAVERMLG	TVIGLGAGLG	VWLNLQHYFH
101	GNLLFYLTVG	TASALAGWAA	VGRNGYVFLM	AGLTMCMLIG	DNGSEWLDSC
151	IMRAMNVLLG	AAIAIAAALK	LPLKSTLMWR	FMLADNLACD	SKMIAEISNC
201	RRMTRERLEE	NMAKMRQINA	RMVKSRSHLA	ATSGESRISP	AMMEAMQHAH
251	RKIVNTTELL	LTTAAKLQSP	KINGSEIRLL	DRHFTLLQTD	LQQTVALINC
301	RHARRIRIDT	AINPELEALA	EHLHYQWQGF	LWLSTNMRQE	ISALVILLQR
351	TRRWLDAHE	RQHLRQSLLE	TRHH*		

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

- 25 ORF146 shows 98.6% identity over a 74aa overlap with an ORF (ORF146a) from strain A of *N. meningitidis*:

	orf146.pep			10	20	30
				RHARRIRIDTAINPELEALAEHLHYQWQGF		
30	orf146a	KINGSEIRLLDRHFTLLQTDLQQTVALINGRARRIRIDTAINPELEALAEHLHYQWQGF				
		280	290	300	310	320
						330
	orf146.pep		40	50	60	70
			LWLSTNMRQEISALVILLQRTRRWLDAHERQHLRQSLLETRHHGX			
35	orf146a		LWLSTNMRQEISALVILLQRTRRWLDAHERQHLRQSLLETRHSHX			
		340	350	360	370	

The complete length ORF146a nucleotide sequence <SEQ ID 633> is:

40	1	ATGAACACCT	CGCAACGCAA	CGCGCTCGTC	AGCCGCTGGC	TCAACTCCTA
	51	CGAACGCTAC	CGCTACCGGC	GCCCTCATCCA	CGCCGCTCGG	CTCGCGGGGG
	101	CGCTCCTGTT	CGCCACCGCC	TCCGCCCGGC	TGCTCCACCT	CCACACACGGC
	151	GAGTGGATAG	GGATGACCGT	CTTCGTGCTC	CTCGGCATGC	TCAGTPTTCA
	201	AGGGCGGATT	TACTCCAGGG	CGGTGGAGCG	TATGCTCGGC	ACGGTCATCG
	251	GGCTGGCGCC	GGGTTTGGCG	GTTTATGCTC	TGAACACAGA	TTATTTCAC
45	301	GGCAACCTCC	TCTTCTACTC	CACCGTGGCC	ACGGCAAGCG	CACCTGGCGCG
	351	CTGGCGGGCG	GTCGGCAAAA	ACGGCTACGT	CCCTATCTGT	GCGGGGCTGA
	401	CGATGTGCAT	GCTCATCGGC	GACAACGGCA	GCGAATGGTT	CGACAGCGGC
	451	CTGATGGCGC	CGATGAACGT	CCTCATCGGC	GCGGCGATCG	CCATCGCCGC
	501	CGCCAAACTG	CTCGCGGTGA	AATCCACACT	GATGTGGAGT	TTCAATGCTTG
50	551	CGGACAACCT	GACCGACTGC	AGCAAAATGA	TTGCCGAAAT	CAGCAACGGC
	601	AGGCGCATGA	CCCGCGAACG	CCTCGAGGAG	AAACATGGGCA	AAATGGGCGCA
	651	AATCAAACGA	CGCATGGTCA	AAAGCCCGAG	CACACTCGCC	GCCACATCGG
	701	GGGAAAGCGG	CATCAGCCCC	GCCATGATGG	AAGCCATGCA	GCAAGCCAC
	751	CGTAAATTCG	TCAACACCAC	CGAGCTGCTC	CTGACCAACG	CGGCCAAGCT
55	801	GCAATCTCCC	AAACTCAACG	GCAGGAAAT	CGGCTGCTT	GACGCCACT
	851	TCACACTGCT	CAACAACGAC	CTGCACAAA	CGCTCGCCCT	TATCAACGGC
	901	AGCACACCTC	GGCGATCCG	CATCGACAC	GCATCAAC	CGCACTCGA
	951	AGCCCTCGCC	GAAACACCTC	ACTTCCAATG	GCAAGGCTTC	CTCTGGCTCA
	1001	GCAACCAATAT	GGCTCAGGAA	ATTTCGCCCT	TGCTCATCTC	GCTGCAACGC
60	1051	ACCCGCGCA	AATGGCTGGA	TGCCCAAGAA	GCACAACCC	TGCGCCAAAG
	1101	CCTGCTTGAA	ACACGGGAAC	ACAGTTGA		

This encodes a protein having amino acid sequence <SEQ ID 634>:

	1	MNTSQNRNLVSRWLSYERYRYRRLIHAVRLGGAVLFATASARLLHLQHGEWIGMTVFVV
	51	EWIGMTVFVV LGMLQFGAI YSKAVERMLG TVIGLGAGLG VLWLNQHYFH
	101	GNLLFYLTVG TASALAGWAA VGKNGYVPM LAGLTMCMLIG DNGSEWFDG
5	151	LMRAMNVLI AAIATAAKL LPLKSTLMWR FMLADNLDCS XMIAEISNG
	201	RMTRERLEE NMAKMRQINA RMVKSRLHA ATSGESRISP AMMEAMQHAH
	251	RKIVNTTELL TTAALKQSP KLNGSEIRLL DRHFTLLQTD LQQTVALING
	301	RHARRIRIDT AINPELEALA EHLHYQWQGF LWLSTNMRQE ISALVILLQR
	351	TRRKWLDAHE RQHLRQSLLE TREHS*
10	ORF146a and ORF146-1 show 99.5% identity in 374 aa overlap:	
	orf146a.pep	MNTSQNRNLVSRWLSYERYRYRRLIHAVRLGGAVLFATASARLLHLQHGEWIGMTVFVV
	orf146-1	MNTSQNRNLVSRWLSYERYRYRRLIHAVRLGGAVLFATASARLLHLQHGEWIGMTVFVV
15	orf146a.pep	LGMLQFGAIYSKAVRMLGTVIGLGAGLGVLWLNQHYFHGNLLFYLTVG TASALAGWAA
	orf146-1	LGMLQFGAIYSKAVRMLGTVIGLGAGLGVLWLNQHYFHGNLLFYLTVG TASALAGWAA
	orf146a.pep	VGKNGYVPM LAGLTMCMLIGDNGSEWFDGSLMRAMNVLI GAAIAATAAKL LPLKSTLMWR
20	orf146-1	VGKNGYVPM LAGLTMCMLIGDNGSEWFDGSLMRAMNVLI GAAIAATAAKL LPLKSTLMWR
	orf146a.pep	FMLADNLDCS XMIAEISNGRRMTRERLEENMAKMRQINARMVKSRSHLAATSGESRISP
25	orf146-1	FMLADNLDCS XMIAEISNGRRMTRERLEENMAKMRQINARMVKSRSHLAATSGESRISP
	orf146a.pep	AMMEAMQHARRKIVNTTELLTTAAKLQSPKLNGSEIRLLDRHFTLLQTDLQQTVALING
	orf146-1	AMMEAMQHARRKIVNTTELLTTAAKLQSPKLNGSEIRLLDRHFTLLQTDLQQTVALING
30	orf146a.pep	RHARRIRIDTAINPELEALAEHLHYQWQGF LWLSTNMRQEISALVILLQTRRKWLDAHE
	orf146-1	RHARRIRIDTAINPELEALAEHLHYQWQGF LWLSTNMRQEISALVILLQTRRKWLDAHE
35	orf146a.pep	RQHLRQSLLE TREHSX
	orf146-1	RQHLRQSLLE TREHG

Homology with a predicted ORF from *N.gonorrhoeae*

- 40 ORF146 shows 97.3% identity over a 75aa overlap with a predicted ORF (ORF146ng) from *N.gonorrhoeae*:

	orf146.pep	RHARRIRIDTAINPELEALAEHLHYQWQGF	30
	orf146ng	KLNGSEIRLLDRHFTLLQTDLQQTAAALINGRHARRIRIDTAINPELEALAEHLHYQWQGF	364
45	orf146.pep	LWLSTNMRQEISALVILLQTRRKWLDAHERQHLRQSLLE TREHG	75
	orf146ng	LWLSTNMRQEISALVILLQTRRKWLDAHERQHLRQSLLE TREHG	409

An ORF146ng nucleotide sequence <SEQ ID 635> was predicted to encode a protein having amino acid sequence <SEQ ID 636>:

	1	MSGVRFPSA	PIPSTDPPSG	SLCFTFTPLQ	TASDMNSQQR	KRLSGRWINS
	51	VERYRHRLI	HAVRLGGTVL	FATALARLL	LQHGEWIGMT	VFVVLQMLQF
	101	QGAISNAVE	RMLQTVIGLG	AGLGVLWLNQ	HYFHGNLLFY	LTGTASALA
55	151	GWAAGVKNY	VPM LAGLTMC	MLIGDNGSEW	LDSGLMRAMN	VLIGAAIATA
	201	AAKLLPLKST	LMWRFLADN	LADCSKIMAE	ISNGRRMTR	REONMVKMR
	251	QINARMVKSR	SHLAATSGES	RISPSMMEAM	QHAHRKIVNT	TELLTTAAK
	301	LQSPKLNGSE	IRLLDRHFTL	LQTDLQQTAA	LINGRHARRI	RIDTAINPEL
	351	EALAEHLHYQ	WQGF LWLSTN	MROEISALVI	PLQTRRKWL	DAHERQHLRQ
	401	SLLE TREHG*				

- 60 Further work revealed the following gonococcal DNA sequence <SEQ ID 637>:

1 ATGAACTCCT CGCAACGCCA ACGCCCTTTC GgccGCTGGC TCAACTCCTA
 51 CGAACGCTac cGCGacCGCC GCCTCATACA TGCCGTGCGG CTCGGCGgaa
 101 ccgtCCTGTT CGCCACCGCA CTCGCCCGgc tACTCCACCT CCAAcacgyc
 151 gAATGGATAG GGAatgaCCGT CTCGCTCGTC CTCGGCATGC TCCAGTTCCA
 201 AGGCGcgatt tActccaaag ccggtgGAacg taTgtctcgt acggtcatcg
 251 ggctgGGCGC GGGTTTGGgc gTTTATGGC TGAACCAGCA TTAttccac
 301 ggcacacCTcc tttcttaact gacatcggc acggcaagcg cactggccgg
 351 ctGGCGGGCG GTCCGGCAAAA acgggttaegt ccttatgtcg CGGGGGctgA
 401 CGAATGCTAT gctaatcggc GACAACGGCA GGAATGGCT CGACACGGC
 451 CTGATGCGCG CGATGAACCT CCTCATCGGC CGCGCATCG CCAITGGCGC
 501 CGCCAAACCTG CTGCGCGTGA AATCCACACT GATGTGGCGT TTATGCTGTG
 551 CGCAACACCT GCGCGACTCG AGCAAAATGA TTGCGAAAT CAGCAACGGC
 601 AGGCGTATGA CGCGGCAAGC TTTGGAGCAG AATATGGTCA AATGCGCGCA
 651 AATCAACGCA CGCATGGTCA AAAGCGCGAG CCACCTCGCC GCCACATCGG
 701 GCGAAAGCCG CATCAGCCCC TCCATGATGG AAGCATGCA GACAGCCAC
 751 CGCAAAATCG TCAACACCAC CGAGCTGCTC CTGACCACCG CGCGCAAGCT
 801 GCAATCTCCC AAATCAACG CGAGCGAAAT CGGCTGCTC GACCGCACT
 851 TCACACTGCT CCAACCGGAC CTGCAACAAA CGCGCGCCCT CATCAACGGC
 901 AGACACGCCC GCGCGATCG CATGCAACCC GCCATCAACC CGGAATCGGA
 951 AGCCCTCGCC GAACACCTCC ACTACCAATG CGAGGGCTTC CTCTGGCTCA
 1001 GCACCAATAT GCGTCAAGAA ATTTCCGCCC TCCTGCTCCT GCTGCAACCG
 1051 ACCCGCGCA AATGCGTGGG TGCCACGAAA CGCAACACG CGGCGCAAG
 1101 CCTGCTTGA ACACGGGAAC ACGGCTGA

This corresponds to the amino acid sequence <SEQ ID 638; ORF146ng-1>:

1 MNSSQRKRLS GRWLSNERY RHRLIHAVR LGGTVLFAA LARLLHLQHG
 51 EWIGMTVFVV LGMLQFOGAI YSNAVERMLG TVIGLGAGLG VLWLNQHYHF
 101 GNLLFYLTIG TASALAGWAA VGKNGYVPM LAGLTMCLIG DNGSEWLDG
 151 LMRAMNVLIG AAIATAAKL LPLKSTLMWR FMLADNLAD SKMIAEISNG
 201 RMRTRERLEQ NMVKMRQINA RMVKSRSILA ATSGESRISP SMMEAMQHAH
 251 RKIVNTTELL LTAAKLQSP KLNGSEIRLL DRHFTLLQDT LQQTAAALNG
 301 RHARRIRIDT AINPELEALA EHLHYQWQGF LWLSTNMRQE ISALVILLQR
 351 TRRWLDAHE RQHLRQSLLE TREHG*

ORF146ng-1 and ORF146-1 show 96.5% identity in 375 aa overlap

35 orf146-1.pep MNTSQNRNLSVRWLSNYFRYRRLIHAVRLGGAVLFATASARLLHLQHGWEIGMTVFVV
 orf146ng-1 MNSSQRKRLSGRWLSNERYRHRLIHAVRLGGTVLFATALLHLQHGWEIGMTVFVV
 40 orf146-1.pep LGMLQFOGAIYSKAVRMLGTVIGLGAGLGVLWLNQHYHFGNLLFYLTIGTASALAGWAA
 orf146ng-1 LGMLQFOGAIYSNAVERMLGTVIGLGAGLGVLWLNQHYHFGNLLFYLTIGTASALAGWAA
 45 orf146-1.pep VGKNGYVPM LAGLTMCLIGDNGSEWLDGSLMRAMNVLIGAAIAIAAAKLLPLKSTLMWR
 orf146ng-1 VGKNGYVPM LAGLTMCLIGDNGSEWLDGSLMRAMNVLIGAAIAIAAAKLLPLKSTLMWR
 50 orf146-1.pep FMLADNLADCSKMIAEISNGRMRTRERLEENAKMRQINARMVKSRSILAATSGESRISP
 orf146ng-1 FMLADNLADCSKMIAEISNGRMRTRERLEENAKMRQINARMVKSRSILAATSGESRISP
 55 orf146-1.pep AMMEAMQHAHRKIVNTTELLTTAAKLQSPKLNSEIRLLDRHFTLLQDTLQQTAAALNG
 orf146ng-1 SMMEAMQHAHRKIVNTTELLTTAAKLQSPKLNSEIRLLDRHFTLLQDTLQQTAAALNG
 60 orf146-1.pep RHARRIRIDTAINPELEALAEHLHYQWQGF LWLSTNMRQE ISALVILLQRTTRRWLDAHE
 orf146ng-1 RHARRIRIDTAINPELEALAEHLHYQWQGF LWLSTNMRQE ISALVILLQRTTRRWLDAHE
 orf146-1.pep RQHLRQSLLETREHG
 orf146ng-1 RQHLRQSLLETREHG

Furthermore, ORF146ng-1 shows homology with a hypothetical *E. coli* protein:

sp|P33011|YEAE_ECOLI_HYPOTHETICAL_40.0_KD_PROTEIN_IN_COBU-SMNC_INTERGENIC_REGION
 >gi|1736674|gn|P|D|d|016553 (D90838) ORF_ID:c348#20; similar to [SwissProt
 Accession Number P33011] [Escherichia coli] >gi|1736682|gn|P|D|d|016560 (D90839)
 ORF_ID:c348#20; similar to [SwissProt Accession Number P33011] [Escherichia coli]

-355-

>gi1788318 (AE000292) f352; 100% identical to fragment YESEA_ECOLI SW: P33011 but has 203 additional C-terminal residues [Escherichia coli] Length = 352
Score = 109 bits (271), Expect = 2e-23
Identities = 89/347 (25%), Positives = 150/347 (42%), Gaps = 21/347 (6%)

5 Query: 20 YHRRRIHAVRLGGTVLFATALARLLHLQHGGEWIGMTVFVVLGMLQFQGAISNAVERML 79
YRH R++H R+ L + RL + W +T+ V++G + F G + A ER+
Sbjct: 15 YHRYRIVHGTGRVALAFLTLFLIIRLFTPIESTWPLVTMVINGPISFWGNVVRFAFERIG 74

10 Query: 80 GTVIGLCAGLGVLNHLNQHYFHNLLPYLTIGTASALAGWAAGCKGVPMIAGLTMCMCLI 139
GTV+G GL L L L L + A L GW A+GK Y +L G+T+ +++
Sbjct: 75 GTVIGLSILGIALQLE---LISLPLMLVWCAAMFLOGWALGKKPYQGLLIGVTLAIVV 131

15 Query: 140 GDNGSEWLDGSLMRAMNVLIQXXXXXXXXXXKLPLKSTIMWRFMADNLADCKSMIAEISN 199
G E +D+ L R+ +V++G + P ++ + WR LA +L + +++ +
Sbjct: 132 GSPTGE-IDTALWRSQDVLGSLIAMLFTGIWQORAFTHWRIQLAKSLTEYNRVYQSAPS 190

20 Query: 200 GRMTRERLEQNMVMKMRQINARMVKSRSHLAATSGESRISPSMMEAMQHAHRKIVNXXXX 259
+ R RLE ++ K+ VK R +A S E+RI S+ E +Q +R +V
Sbjct: 191 PNLRFRPLESHLQKLL---TDAVMKRLGIAPASKETRIKPSIYEGITQINRNLVCMLE 247

25 Query: 260 XXXXXXXXQSPK---LNGSEIRLLDRHFXXXXXXXXXXAAALINGHARRIRIDTAINPEL 316
+ LN ++R D + AL G +N +
Sbjct: 248 QINAYWATRPSHFVLLNQAQLR---DTQHMMQQTLLSLVHALYEGNPQPVFANTEKLDNAV 305

Query: 317 EALAEHL--HYWQ-----GFLVSTNMQRQISALVILLQRTRRK 354
E L + L H+ + G++WL+ ++ L L+ R RK
Sbjct: 306 EELRQLNNHHLKVVETPIYGYVWLNMETAHQLELLSNLICRALRK 352

On the basis of this analysis, including the identification of several transmembrane domains in the
30 gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 76

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 639>

1 ..CCGCAACACA CGCGCGTTAC CGCACAGCTT TTGAGCCCGT ACGGGATTCA
35 51 CGCGCAAACTC GTCACTGTGC CGCAACACRA CGAAGCGCAG ATGGCGGACA
101 AGATTGTGGC CTATCTTTCA CAGCGCATGG TTGTGCGCAT GGTTCGCAT
151 CGCGGTACCG CGCGCGTGTG CGACCGCGGC CGCAAACTCG CCGCGCGCGT
201 CGGTGAGGCG GGGTTTAAAG TCGTTCCCGT CGTGGCGCA AC.GCGGTGA
40 251 TGCGCGCTTT GAGCGTGCCG GGTGTGGAAG GATCCGATT TTATTTCAC
301 GGTTTGTATC CGCGCAAAATC GGGAGAAGCG AGGAACACTG TTGCAAAAT
351 GGTGCGGCGC GCGTTTCTCA TCGTCATGTT TGAACGCGCG CACCGCATCG
401 GTGCAGCGCT TGCGGATATG CGGGAAGTGT TCCTCGAAGC CGAATTAATG
451 CTGCGCGCGC AATTAACGAA AACGTTTGAA ACGTTCTTAA CGGCAACGGT
501 TGCGGGAATTT CAGACGGCAT TGCTGCGCA CGCGGACCAA TCGCGCGCGC
45 551 AGATGGTGTG GGTGCTTTAT CGGCGCGCAG ATGAAAAACA CGAAGCGTTG
601 TCGAGTCCG CGCAAAACAT CATGAAATC CTCACGCGCG AGCTGCGCAG
651 CAACGAGCGC GCGGAGCTTG CTGCAAAAT CACGGCGCAG GGAAGAGAG
701 CTTTGTACGA T..

This corresponds to the amino acid sequence <SEQ ID 640; ORF147>:

50 1 ..AEDTRVTAQL LSAYGIOGKL VSVREHNERQ MADKIVGYLS DGMVVAQVSD
51 AGTPAVCDPG AKLARRVREA FKVVPVPGA XAVMAALSA GVEGSDPFRN
101 GFVPPKSGER RKLPAKWRA AFPIVMFETP HRIGALADM AELFPERRIM
151 LAREITKTFE TPLSGTVGEI OTALSDAGDQ SRGEMVLVLY PAQDEKHEGL
201 SESAQNMIMKI LTAEPLTKQA AELAAKITGE GKALYD..

55 Further work revealed the complete nucleotide sequence <SEQ ID 641>:

1 ATGTTTCAGA AACATTGCA GAAAGCCTCC GACAGCGTCG TCGAGGGGAC
51 ATTATACGTG GTTGGCAGCG CCATCGGCAG TTTGGCGGAC ATTACCTCGT
101 CGGCTTTGGC GGTATTGCAA AAGGCGGACA TCATCTGTGC CGAAGACACG
151 CGGCTTACCG CACAGCTTTT GAGCGCGTAC GGCATTGAG GCACAACTCGT

201 CAGTGTGCGC GAACACAACG AACGGCAGAT GCGGACAAAG ATTGTGCGCT
 251 ATCTTTTCAGA CGCGATGTTT GTGGCAGCAGG TTCCGATGCG GGGTACGCG
 301 GCGGTGTGCG ACCCGGCGCG GAACTCGCC CGCGCGGTGC GTGAGGCGCG
 351 GTTTAAAGTC GTTCCCGTCG TGGCGCGCAAG CGCGGTGATG GCGGCTTTGA
 401 GCGTGGCGCG TGTGGAAGGA TCCGATTTTT ATTTCAACGG TTTGTACCG
 451 CCGAAATCGG GAGAACGCGAG GAACTGTTT GCCAATGGG TCGCGGCGCG
 501 GTTTCCTATC GTCATGTTTG AAACGCGCGA CGCGATCGGT GCGACGCTTG
 551 CCGATATGGC GGAATCTGTT CCGACGCGCG GATTATATGCT GCGCGCGGAA
 601 ATTACGAAGA GTTTTGAAC GTTCTTAAGC GCGACGCTTG GCGAAATCTG
 651 GACGCGATCG TCTCCGACG GGAACCAATC GCGCGCGAG ATGCTTTTGA
 701 TGCTTTATCC GCGCGAGGAT GAAACAACG AGAGCTTTGC CGAGTCGCGC
 751 CAACAACATCA TGAAATCTC CACAGCGCAG CTGCGCGAGC AACAGGCGCG
 801 GGAGCTTGCT GCCAAATCA CCGGCGAGGG AAGAGAAAGCT TTGTACGATC
 851 TGGCTCTGTC TTGGAAGAAC AAATAG

15 This corresponds to the amino acid sequence <SEQ ID 642; ORF147-1>:

1 MFQKHLQKAS DSVVGGTLYV VATPIGNLAD ITRALAVLQ KADIICAEDT
 51 RVTAQLLSAY GIQKLVSVR EHNERQMAK IVGYLSDGMV VAQVSDAGTP + DPG
 101 AVCDPGAKLA RRVREAGFKV VEVVGASAVM AALSVAGVEG SDFYFNGFVP
 151 PKSGERRKLF AKWVRAAFPI VMFETPHRIG ATLADMAELF PERRIMLARE
 201 ITKTFETFLS GTVGEIQTALSADGQSRGEMVLVLPQAQ DEKHEGLSESA
 251 QNIMKILTAE LPTKQAAELA AKITGEGKKA LYDLALSWMN K*

Computer analysis of this amino acid sequence gave the following results:

Homology with hypothetical protein ORF286 of *E. coli* (accession number U18997)

ORF147 and *E. coli* ORF286 protein show 36% aa identity in 237aa overlap:

25 Orf147: 1 AEDTRVTAQLLSAYGIQKLVSVREHNERQMAKIVGYLSDGMVVAQVSDAGTPAVCDPG 60
 AEDTR T LL +GI +L ++ +HNE+Q A ++ L +G +A VSDAGTP + DPG
 Orf286: 43 AEDTRHTGLLLQHFGINARLFALHDHNEQQAETLLAKLEGQONIALVSDAGTFLINDPG 102
 30 Orf147: 61 AKLARRVREXXXXXXXXXXXXXXXXXEGSDYFNGFVPPKSGERRKLFKAVRA 120
 L R RE F + GF+P KS RR
 Orf286: 103 YHLVRTCREAGIRVVPLPGPCAATIALSAAGLSPDRFCYEGFLPAKSKGRRLDALKAEAE 162
 35 Orf147: 121 AFPIVMFETPHRIGALADMAELFPERR-LMLAREITKTFETFLSGTVGEIQTALSADG 179
 ++ +E+ HR+ +L D+ + E R ++LARE+TKT+ET VGE+ + D +
 Orf286: 163 PRTLFIYESTHRLDLSIEDIVAVLGSERYVVLARELTKTWETIHGAPVGEALLAWVKEDEN 222
 40 Orf147: 180 QSRGEMVLVLYPAQDEKHEGLSESAQNIMKILTAEPLTKQAAELAAKITGEGKKALY 236
 + +GEMVL++ + E L A + +L AELP K+AA LAA+I G K ALY
 Orf286: 223 RRGEMVLV-EGHKAQEDLPADALRLTALLQAEPLKKAALAAEIHGVKKNALY 278

Homology with a predicted ORF from *N. meningitidis* (strain A)

ORF147 shows 96.6% identity over a 237aa overlap with ORF75a from strain A of *N. meningitidis*:

45 orf147.pep AEDTRVTAQLLSAYGIQKLVSVREHNERQ 10 20 30
 orf75a TLYVVATPIGNLADITRALAVLQKADIICAEDTRVTAQLLSAYGIQKLVSVREHNERQ
 20 30 40 50 60 70
 50 orf147.pep MADKIVGYLSDGMVVAQVSDAGTPAVCDPGAKLARRVREAGFKVVPVVGASAVMAALSV 40 50 60 70 80 90
 orf75a MADKIVGYLSDGMVVAQVSDAGTPAVCDPGAKLARRVREAGFKVVPVVGASAVMAALSV 80 90 100 110 120 130
 55 orf147.pep GVEGSDFYFNGFVPPKSGERRKLFKAVWRAAFPIVMFETPHRIGALADMAELFPERRIM 100 110 120 130 140 150
 orf75a GVAGSDFYFNGFVPPKSGERRKLFKAVWRAAFPIVMFETPHRIGATLADMAELFPERRIM 140 150 160 170 180 190
 60 orf147.pep LAREITKTFETFLSGTVGEIQTALSADGQSRGEMVLVLYPAQDEKHEGLSESAQNIMKI 160 170 180 190 200 210

	orf75a	 LAREITKTFETFLSGTVGEIQTALAADGNQSRGEMVLVLYPAQDEKHEGLSESAQNIMKI 200 210 220 230 240 250
5	orf147.pep	220 230 LTAELPTKQAAELAAKITGEGKKALYD
	orf75a	LTAELPTKQAAELAAKITGEGKKALYDLALSWKNKX 260 270 280 290
10	ORF147a is identical to ORF75a, which includes aa 56-292 of ORF75.	
	<u>Homology with a predicted ORF from <i>N. gonorrhoeae</i></u>	
	ORF147 shows 94.1% identity over a 237aa overlap with a predicted ORF (ORF147ng) from <i>N. gonorrhoeae</i> :	
15	orf147.pep	AEDTRVTAQLLSAYGIQGLKLVSVREHNERQ 30
	orf147ng	TLYVVATPIGNLADITLRALAVLQKADIIAEDTRVTAQLLSAYGIQGLKLVSVREHNERQ 85
	orf147.pep	MADKIVGVLSDGMVVAQVSDAGTAPVCDPGAKLARRVREAGFKVVPVVGASAVMAALSVA 90
20	orf147ng	MADKIVGVLSDGLVVAQVSDAGTAPVCDPGAKLARRVREAGFKVVPVVGASAVMAALSVA 145
	orf147.pep	GVEGSDFYFNGFVPPKSGERRKLFAKWVRAAFPIVMFETPHRIGALADMAELFPERRLM 150
25	orf147ng	GVAESDFYFNGFVPPKSGERRKLFAKWVRAAFPVVMFETPHRIGATLADMAELFPERRLM 205
	orf147.pep	LAREITKTFETFLSGTVGEIQTALSDAGDQSRGEMVLVLYPAQDEKHEGLSESAQNIMKI 210
	orf147ng	LAREITKTFETFLSGTVGEIQTALAADGNQSRGEMVLVLYPAQDEKHEGLSESAQNIMKI 265
30	orf147.pep	LTAELPTKQAAELAAKITGEGKKALYD 237
	orf147ng	LAAELPTKQAAELAAKITGEGKKALYDLALSWKNK 300

An ORF147ng nucleotide sequence <SEQ ID 643> was predicted to encode a protein having amino acid sequence <SEQ ID 644>:

35	1 MSVFQTAFFM FQKHLQKASD SVVGGTLYVV ATPIGNLADI TLRALAVLQK 51 ADIIAEDTR VTAQLLSAYG IQGLRVSVRE HNERQMAQDV IGFLSDGLVV 101 AOVSDAGTFA VCDPGAKLAR VRREAGFKV PVVGASAVMA ALSVAGVAES 151 DFYFNGFVPP KSGERRKLFA KKWVRAAFV FMEFTHRIGA TLADMAELFP 201 ERLRLMLAREI TKTFTETLSG TVGEIQTALA ADGNQSRGEM VLVLYPAQDE 40 251 KHEGLSESAQ NAMKILAAEL PTKQAAELAA KITGEGKKAL YDLALSWKNK 301 *
----	---

Further work revealed the following gonococcal DNA sequence <SEQ ID 645>:

	1 ATGTTTCAGA AACACTTGCA GAAAGCTCC GACACGCTCG TCGGAGGGAC 51 ATTATACGTG GTTGCCAGCG CCACTCTGCA ATTACCTGCA 45 101 GCGCTTTGCG GGTATTGCGA AAGCGGGACA TCATTCTGCG CGAGACACG 151 CGCGTTACTG CCGAGCTTTT GAGCGCGTAC GGCATTCAAG CAGGTTGGT 201 CAGTGTGCGC GAACACACCG AGCGCGCATG GCGCGACAAG GTAATCGGTT 251 TCCTTTCAGA CGGCTGTGTT GTGGCGCAGG TTTTCGATGC GGGTACGCCG 50 301 GCGCTGTGCG ACCCGGCGCG GAAACTCGCC CGCGCGCTCG GCGAAGCAGG 351 GTTCAAAGTC GTTCCCGTCG TGGCGCGAAG CGCGGTAATG CGGCGGTTGA 401 GTGTGGCCGG GTGTGGCGAA TCGGATTTTT ATTTCAACGG TTTTGTACCG 451 CGGAATTCGG GCGAACGCTAG GAAATTTGTT GCCAATGGG TCGGGCGCGC 501 ATTTCTGTGC GTCATGTTTG AAACCGCGCA CCGAATCGGG GCAACGCTTG 55 551 CCGATATGCG GGAATTGTTT CCGGAACGCC GTCTGATGCT GCGCGCGGAA 601 ATCAACGAAA CGTTTGAACG GTTCTTAAGC GGCACGGTTG GGGAAATTCA 651 GACGGCATTG GCGCGCGAGC GCACACCAAT CGCGCGCGAG ATGGTGTGTTG 701 TGTCTTATCC GCGCGAGCAT GAATAACACG AAGGCTGTGC CGAGTCTGCG 751 CAAATTCGCA TGAATATCTT TCGCGCGCAG CTGCGACACA AGCAGCGCGC 801 GGAGCTTGCC GCCAAGATTA CAGTGAGGGC CAAAAGGCTT TTGTACGATT 60 851 TGGCACTGTC GTGAAAAAC AATATGA
--	--

This corresponds to the amino acid sequence <SEQ ID 646; ORF147ng-1>:

```

1  MPQKHQKAS  DSVVGGTLYV  VATPIGNLAD  ITLRALAVLQ  KADICAEDT
51 RVTAQLLSAY  GIQGRILSVR  EHNERQMAK  VIGFLSDGLV  VAQVSDAGTP
101 AVCDPGAKLA  HRVREAGFKV  VEVVGASAVM  AALSVAAGV  SDFYFNGFVP
151 PKSGERRKLF  AKWVRAAFV  VMFETPHRIG  ATLADMAELF  PERRIMLARE
201 ITKTETFTFLS  GTVGEIQTAL  AADGNQSRGE  MVLVLVPAQD  EKHEGLSESA
251 QNAMKILAAE  LPTKQAAELA  AKITGEGKKA  LYDLALSWKN  K*

ORF147ng shows homology to a hypothetical E. coli protein:

10  sp|P45528|YRAL_ECOLI  HYPOTHETICAL 31.3 KD PROTEIN IN AGAI-MTR INTERGENIC REGION
    (F286)
    >gi|606086 (U18997) ORF_f286 [Escherichia coli]
    >gi|1789535 (AE000395) Hypothetical 31.3 kD protein in agai-mtr intergenic region
    [Escherichia coli] Length = 286
    Score = 218 bits (550), Expect = 3e-56
    Identities = 128/284 (45%), Positives = 171/284 (60%), Gaps = 4/284 (1%)

15  Query: 4  KHLQKASDSVVGGLTYVATPIGNLADITLRLAVLQKADICAEDTRVTAQLLSAYGQI 63
      K Q A +S G LY+V TPIGNLADIT RAL VLQ D+I AEDTR T LL +GI
      Sbjct: 2  KQHGSADNSQ--GQLYIVPTPIGNLADITQRALEVLQAVDLIAEDTRTGLLLQHFGIN 59

20  Query: 64  GRLVSVREHNERQMAKDVIGFLSDGLVVAQVSDAGTPAVCDPGAKLARLVREAGFKVVPV 123
      RL ++ +HNE+Q A+ ++ L +G +A VSDAGTP + DPG L R REAG +VVF+
      Sbjct: 60  ARLFALHDHNEQQKAETLLAKLQEQNTALVSDAGTFLINDPGYHLVTRCREAGIRVVP 119

25  Query: 124  VGASAVMAALSVAAGVSDFYFNGFVPPKSGERRKLFKAVVRAAFVVMFETPHRIGATL 183
      G A + ALS AG+ F + GF+P KS RR ++ +E+ HR+ +L
      Sbjct: 120  FPGCAAITALSAAGLPSDRFCYEGFLPAKSKGRRDALKATIEAEPRTLIFYESTHRLDLSL 179

30  Query: 184  ADMALFPER--LMLAREITKTFTFLSTVGEIQTALAADGNQSRGEMVLVLVPAQDEK 242
      D+ + ER ++LARE+TKT+ET VGE+ + D N+ +GEMVL++ +
      Sbjct: 180  EDIVAVLGESRYVVLARELTKTWETHGPAVGEGLAWVKEDENRRKEMVLVIV-EGHKAQ 238

35  Query: 243  HEGLSESAQNAKMLIAELPTKQAAELAAKITGEGKKALYDLAL 286
      E L A + +L AELP K+AA LAA+I G K ALY AL
      Sbjct: 239  EEDLPADALRTLAILQAEPLKKAALAAEIHGVKKNALYKYAL 282
  
```

Based on the computer analysis and the presence of a putative transmembrane domain in the gonococcal protein, it is predicted that these proteins from *N. meningitidis* and *N. gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 77

40 The following partial DNA sequence was identified in *N. meningitidis* <SEQ ID 647>

```

1  ATGAAACAA  CCGACAAACG  GACAAACGAA  ACACACCGCA  AAGCCCCGAA
51  AACCGGTCGC  ATCCGCTTCT  C. GCTGCCTTA  CTTAGCCATA  TGCCTGCGT
101  TCGCGATTCT  TCCCCAGCCG  TGGCGCGGAC  ACACATTATT  CGGCATCAAC
151  TACCAATACT  ATCGCGACTT  TGCCGAAAT  AAGGCAAGT  TTGCAGTCGG
45  201  GCGCAAGAT  ATTGAGGTTT  ACACACAAAA  AGGGAGTTG  TFCGCGAAAT
251  CAATGACAAA  AGCCCCGATG  ATTGATTTTT  CTGTGTGTCT  GCGTAAACGG
301  GTGGCGGCGAT  TGGTGGCGGT  ATCAATATAT  TGTGAGCGTG  GCACATAACG
351  CGCGCTATAA  CAACGTTGAT  TTTGTGCGCG  AAGGAAG. AA  TATCCC. GAT
401  CAACACGWW  TTACTATATA  AATTGTGAAA  CGGAATAAAT  ATAAAGCAGG
50  451  GACTAAGGC  CATCCTTATG  GCGCGGATTA  TCATATGCGC  CGTTTGCATA
551  AATWGTGCAC  AGATGCAGAA  CCTGTTGAAA  TGACCAGTTA  TATGATGGG
551  CGGAATATA  TCGATCAAAA  TAATTACCTT  GACCGTGTTC  GTATTGGGGG
601  AGCGAGGCAA  TATGGCGCAT  CTGATGAAGA  TGAGCCCAAT  AACCDCGAAA
651  GTTCATATCA  TATTGCAAGT  .....
55  701  .....  GGCTC  ACCAATGTTT  ATCTATGATG  CCGAAAGCA
751  AAATGGGTTA  ATTAATGGGG  TATTGCAAC  GGGCAACCCC  TATATAGGAA
801  AAAGCAATGG  CTTCCAGCTG  GTTCGTAAG  ATTTGTTCTA  TGATGAAATC
851  TTTGCTGGAG  ATACCCATTC  AGTATTCTAC  GAACACGTC  AAAATGGGAA
901  ATACTCTTTT  AACGACGATA  ATAATGGCAC  AGSAAAAATC  AATGCCAAAC
  
```

951	ATGAACACAA	TTCTCTGCCT	AATAGATTAA	AAACACGAAC	CGTTCAATTG
1001	TTTAATGTTT	CTTTATCCGA	GACGACGAAG	GACCTGTGTT	ATCATGTGTC
1051	AGGTGGTGTC	AACAGTTATC	GACCCAGACT	GAATAATGGA	GAATAATATT
1101	CTTTTATTGA	CGAAGGAAAA	GGCGAATTGA	TACTTACGAC	CAACATCAAT
1151	CAAGGTGCTG	GAGGATTATA	TTTCCAAGGA	GATTTTACGG	TCTCGCCTGA
1201	AAATAACGAA	ACTTGGCAAG	GGCGGGGGGT	TCATATCAGT	GAAGACAGTA
1251	CGTTACTCTG	GAAGATAAAC	GGGTGGGCAA	ACGACCGCCT	GTCCAAAATC
1301	GGCAAGGGCA	CGCTG.....
10	2101CATAAAG
2151	TGACTGCTCT	ATTGACTAAG	ACCGACATCA	GGCGCAATGT	CGATCTTGCC
2201	GATCACGCTC	ATTTAAATCT	CACAGGGCTT	GCCCACTCA	ACGCCAATCT
2251	TAGTGCAAAAT	GGCGATACAC	GTTTATCAGT	CGCCACCAAC	GCCACCCAAA
2301	ACGCGCAACCK	TAqCCTCGTG	G.sAATG-CC	AAGCAACATT	TAATCAAGCC
2351	ACATTAAACG	GCAACACATC	GGCTTCgGAG	AATGCTTCAT	TTAATCTAAG
2401	CGACACGCGC	GTACAAAACG	GCAGTCTGAC	GCTTTCGCGC	AACGCTAAGG
2451	CAAAAGTAAAG	CCATTCCGCA	CTCAACGGTA	ATGTCTCCCT	AGCCGATAAG
2501	GCAGTATTCC	ATTTTGAAG	CAGCGCCTTT	ACCGGACAAA	TCAGCGCGCG
2551	CAagGATACG	GCATTACACT	TAAAGACAG	CGAATGGAGC	CTGCGCTCAg
2601	GarCGGAATT	AGGCAATTTA	AACCTTGACA	AGGCCACACT	TACACTCAAT
2651	TCGCGCTTAT	GCCACGATCG	GGCAGGGGCG	CAACCGCGCA	GTGCGACAGA
2701	TGCGCGCGCG	CGCGCTTGCG	CGCGTTCGCG	CGCGTTCGCTA	TTATTCGCTTA
2751	CACCGCCAAC	TTCCGCTAGAA	TCCCGTTTCA	ACACGCTGAC	GGTAAACGGC
2801	AAATTGAACG	GTACAGGAAAC	ATTCCGCTTT	ATGTGGAAGC	TCTTCGCGTA
2851	CCGACAGGAC	AAATTGAAGC	TGGCGGAAAG	TTCCGAAGCG	ACTTACACCT
2901	TGGCGGTCAA	CAATACCGCG	AACGAACCTG	CAAGCTTCGA	ACAATTGACG
2951	GTAGTGGAAAG	GAAGAGACAA	CAACACCGCTG	TCGGAACAC	TTAATTTCAC
3001	CTGCAAAAC	GAACACGCTG	ATGACGGGCG	GTGG.....
30	3551TTAGAC	CGGATATTGG	CGGAAGACCG
3601	CGCAACGCGC	GTTTGGACAA	GGGGCATCOG	GGGACACAAA	CACCTACGGTT
3651	CGCAAGATTT	CGCGCGCTAC	CGGCAACAAA	CGGACCTGCG	CCAAATCGGT
3701	ATGCAGAAAA	ACCTTCGGCAG	CGGGCGGCTC	GGCATTCGTT	TTTTCGCAAA
3751	CGGACGCGAA	AACACCTTCG	ACGAGCGCAT	CGGACATCG	GCACGCGTTG
3801	CCACGCGCGC	CGTTTTCGGG	CAATAGGCGA	TCGACAGGTT	CTCATTCGCG
3851	ATCAGgCGCG	GGCGGGGTTT	TTAGCAGCGCG	CAGCGTTTCA	GACGCGATCG
3901	GAGSMAAAwT	CGCGCGCGCG	GTGCTGCATT	ACGCAATCCA	GGCACGATAC
3951	CGCGCGCGgtt	TCggCGgATT	CGGCATCGAA	CGGCACATCG	GGCGAACCGg
4001	ctATTTCGCTC	CAAAAAGGCG	ATTACCGGTA	CGAAAACGTC	AATATCGCGA
4051	CCCCGCGCCT	TGCATTCAAC	CGCTACGCGG	CGGGCATTTAA	GGCAGATTAT
4101	TCATTCAAAAC	CGCGCGAACA	CATTTCCTAT	ACGCTTATT	TGAGCTGTCT
4151	CTATACCGAT	CGCGCTTCGG	GCAAAAGTCG	AACACGGGTC	AATACCGCGC
4201	TATTGGCTCA	GGATTTCGGC	AAAAACCGCA	GTGCGGAATG	GggCGTAAAC
4251	GGCGAAATCA	AAGGTTTCAC	CGTGTCCCTC	CACGCTGCGC	CGGCCAAAGG
4301	CCGCAACATG	GAAAGCGAAC	ACAGCGGGGG	CATCAAAATTA	GGCTACGCGT
4351	GGTAA...
45	4351TTAGAC	CGGATATTGG	CGGAAGACCG
30	3601	CGCAACGCGC	GTTTGGACAA	GGGGCATCOG	GGGACACAAA
3651	CGCAAGATTT	CGCGCGCTAC	CGGCAACAAA	CGGACCTGCG	CCAAATCGGT
3701	ATGCAGAAAA	ACCTTCGGCAG	CGGGCGGCTC	GGCATTCGTT	TTTTCGCAAA
3751	CGGACGCGAA	AACACCTTCG	ACGAGCGCAT	CGGACATCG	GCACGCGTTG
3801	CCACGCGCGC	CGTTTTCGGG	CAATAGGCGA	TCGACAGGTT	CTCATTCGCG
3851	ATCAGgCGCG	GGCGGGGTTT	TTAGCAGCGCG	CAGCGTTTCA	GACGCGATCG
3901	GAGSMAAAwT	CGCGCGCGCG	GTGCTGCATT	ACGCAATCCA	GGCACGATAC
3951	CGCGCGCGgtt	TCggCGgATT	CGGCATCGAA	CGGCACATCG	GGCGAACCGg
4001	ctATTTCGCTC	CAAAAAGGCG	ATTACCGGTA	CGAAAACGTC	AATATCGCGA
4051	CCCCGCGCCT	TGCATTCAAC	CGCTACGCGG	CGGGCATTTAA	GGCAGATTAT
4101	TCATTCAAAAC	CGCGCGAACA	CATTTCCTAT	ACGCTTATT	TGAGCTGTCT
4151	CTATACCGAT	CGCGCTTCGG	GCAAAAGTCG	AACACGGGTC	AATACCGCGC
4201	TATTGGCTCA	GGATTTCGGC	AAAAACCGCA	GTGCGGAATG	GggCGTAAAC
4251	GGCGAAATCA	AAGGTTTCAC	CGTGTCCCTC	CACGCTGCGC	CGGCCAAAGG
4301	CCGCAACATG	GAAAGCGAAC	ACAGCGGGGG	CATCAAAATTA	GGCTACGCGT
4351	GGTAA...

This corresponds to the amino acid sequence <SEQ ID 648; ORF1>:

1	MKTDKRTTE	THRKAPKTGR	IRFXAAIYAI	CLSPGILPQA	WAGHTYFGIN
51	YQYRYDFEEN	KGKFAVGAKD	IEVYNKGEL	VGKSMTKARM	IDFSVVSRRM
101	VAALVGVQYI	VSVAHNGGYN	MVDFGAEKN	IXDOXRYIVK	TVKRNRYKAG
151	TKGHPYGGDY	HMPLRLHKXT	DAEPVEMTSY	MDGRKYIDQN	NYPRDVRIGA
201	GRQYWRSDSD	EPNRRESSYH	IAS.....GS	PMFIYDAQOG
251	KWLINGVLQT	GNPYIGKSGN	FQLVRKDWFY	DEIFAGDTHS	VFYEPQRNGK
301	YSFNDDNNGT	GKINAKHEHN	SLPNRLKTRT	VQLFNVSLSL	TAREPVYHAA
351	GGVNSYRPLR	NGNENISDFI	EGKGELILTS	NINQAGGLY	FGQDFTVSP
401	NNETWQAGAV	HISEDSTVTW	KVNGVANDRL	SKIKGDTL..
701DKVTAS	LTKTDSIGNV	DLADHAHLNL	TGLATLNGNL
751	SANGDPTTYT	SHNATCNGK	SLVXNAQATF	NQATLNGHTS	ASNGASFNLS
801	DAYVNGSLT	LSQNAKAVLS	HSJLNGVSL	AGDAVFHFES	SRTFGQISGG
851	KDTALHLKDS	EWTLPSGKEI	GNLNLDACT	TLNAGVYRHA	AGATGSAATD
901	APRRSRRSR	RSLLXVTPPT	VSVESFNWLT	VNGKLNGQGT	FRFMSLFGY
951	RSCLKLAEAS	SEGTYYTLAVN	MTNEFPASLE	QLTVVEGKDN	KELSENINFT
1001	LQNEHYDVAGA	W.....
65	1151LDRVFAEDR
1201	RNAVWTSGR	DTMKHYRSQDF	RAYROOTDLR	OIGMQRNLGS	GRVGILFSHN
1251	RTENTFDDGI	GNLSARLAHGA	VFGQYIDIRF	YIGISAGAGF	SSGSLSDGIG
1301	XXKRRRVLYH	GIQARYRAGF	GGFIEPHIG	ATRYFVOKAD	YRYENVNIAI
1351	PLGAFNRYRA	GIKADYSFKP	AQIHISTPYL	SLSYTDAASG	KVTRTRWNTAV

1401 LAQDFGKTRS AEWGVNAEIK GFTLSLHAAA AKGPFQLEAQH SAGIKLYGRW
1451 *

Further sequencing analysis revealed the complete nucleotide sequence <SEQ ID 649>:

5 1 ATGAAACAA CCGACAAACG GACAAACGAA ACACACCGCA AAGCCCGAA
51 AACCGCGCGC ATCCGCTCTC CGCTGCTTA CTTAGCCATA TGCCCTCGT
101 TCGGCATTCT TCCCAAGGCC TGGCGGGGAC ACACTTATTT CGGCATCAAC
151 TACCAATACT ATTGCGACTT TGGCGAAAAT AAAGCGAAGT TTGCGATCGG
201 GGCGAAAGAT ATTGAGGTTT ACACAAAAAA AGGGGAGTTG GTCCGCAAT
10 251 CAATGACAAA AGCCCGGATG ATTGATTTTT GTTGTGTGTC CGGTAGATCG
301 GTGCGGCAAT TCGGCGGCGT CCAATCTGAG GTGAGCGTGG CACATAACGG
351 CGGCTATAAC AACGTGATTT TTGTGTGCGA AGGAAGAAAT CCGCATCAAC
401 ATCGTTTTAC TTATAAAAT GTGAAACGGA ATAATTATA AGCAGGACT
451 AAAGGCCATC CTATTGGCGG CGATTATCAT ATGCGCGGTT TGCATAAATT
501 TGTACAGAT CGAGAACCTG TTGAAATGAC CAGTTATATG GATGGGCGGA
15 551 AATATATCGA TCAAAATAAT TACCCTGACC GTGTTGATG TGGGCGAGCG
601 AGGCAATATT GCGGATCTGA TGAAGATGAG CCAATAACCC GCGAAAGTTC
651 ATATCATATT GCAAGTGGT ATTCTTGGCT CGTTGTGGCG AATACCTTTG
701 CACAAAATGG ATCAGGTGGT GGCAAGTCA ACTTAGGTAG TGAATAAATT
751 AAACATAGCC CATATGGTTT TTATCAACCA GGAGGCTCAT TTGGCGACAG
20 801 TGGCTCACCA ATGTTTATCT ATGATGCCCA AAAGCAAAG TGGTTAATT
851 ATGGGATATT GCAACCGGCG AACCCCTATA TAGGAAAAAG CAATGGCTTC
901 CAGCTGGTTC GTAAAGCTTG GTTCTATGAT GAAATCTTTG CTGGAGATAC
951 CCAATCAGAT TCTTACGAGT CAGCTCAAAA TGGGAATAAC TCTTTTAAAG
1001 ACGATTATAA TGGCAGACGA AAATCAATGT CCAACATGTA ACACATTTCT
25 1051 CTGCTTATTA GATTAAAAAT ACGAACGCTT CAAATTGTTA ATGTTTCTTT
1101 ATCCGAGACA GCAAGAGAAC CTGTTATACA TGCTGCAGGT GGTGTCAACA
1151 GTTATCGACC CAGACTGAAAT AATGGAGAAA ATATTTCCTT TATTGAOGAA
1201 GGAAAGGCGC AATGTATACT TACCAGCAAC ATCAATCAAG GTGCTGAGG
1251 ATTTATTTTC CAAGGAGATT TTACGCTCTC GCCTGAAATT AACGAACTT
30 1301 GGCAAGGCGC GGGCGTTCAT ATCAGTGAAG ACAGTACGCT TACTTGGAAA
1351 GTAAACGCGC TGGCAACAAG CCGCTGTGCC AAAATCGGCA AAGGCAACGT
1401 GCACGTTCAA GCCAAGGGGG AAACCAAGG CTGATCAGC GTGGCGAGCG
1451 GTACAGTCAT TTTGATCAG CAGGCGAGCG ATAAAGGCAA AAAACAAGCC
1501 TTTAGTGAHA TCGGCTTGCT GCGGCGGAGG GGTAGCGTGC AACTGATGTC
35 1551 CGATACACAG GTTAAAGCTTG ACATACTGAT TTTCCGTTCT CGCGCGGACG
1601 GTTGTGATTT AAACGGGCTAT CAGCTTTCGT TCGACGCGAT TCAATAATCG
1651 GATGAAGGGG CGATGATTTG CAACCAAAAT CAAACAAG AATCAACGCT
1701 TACCATTACA GGCATAAAG ATATTCTCTAC AACCGCAAT AACCAACGCT
40 1751 TGGATAGCAA AAAAGAAATT CCTCAACAG GTTGGTTTGG CGAGAAAGAT
1801 ACGACCAAAA GAAACGGGCG GCTCAACCTT GTTTACCGAC CGCGCGAGA
1851 AGACCGCACC CTGCTGCTTT CCGCGCGAAC AAATTTAAC GGCAACATCA
1901 CGCAAAACAA CGCAAACTC TTTTTCAGCG GCAGACCAAC ACGCAAGCC
1951 TACAATCAAT TAAACGACA TTGTCGCAAA AAGAGGGCA TTCTCTCGCG
2001 CGAAATCGTG TGGGCAACAC ACTGATCAA CGGCACATTT AAAGCGGAAA
45 2051 ACTTCCAAAT TAAAGCGGA CAGGCGGTGG TTTCCGCAA TGTTCGCAA
2101 GTGAAAGGCG ATTCGCAATT GAGCAATCAC GCCCAAGCAG TTTTGTGTGT
2151 CGCACCGCAT CAAAGCCACG CAATCTGTAC ACGTGTGAC TGAAGCGGTC
2201 TGACAAATTC TGTCTGAAA ACATTAACCG ACGATTAAGT GATTCCTTCA
2251 TTGACTAAGA CGCAATCAG CGGCAATGTC GATCTGCGG ATACGCTCTA
50 2301 TTTAATCTCT ACAGGGCTTG CCAACTCAA CGGCAATTCT GTTGCATAAT
2351 GCGATACAG TTATACAGTC AGCCACAAG CCACCAAAA CGGCAACCTT
2401 AGCCTGTGCG GCAATGCCA AGCAACATTT AATCAAGCCA CATTAACCGG
2451 CAACACATCG GCTTGGGCA ATGCTTCATT TAATCTAAGC GACCAAGCGG
55 2501 TACAAAACGG CAGTCTGAAG CTCTCGGCA ACGCTAAGG AAACGTAAAG
2551 CATTCGCGAC TCAAGGOTAA TGTCTCCCTA GCGGATAAGC CAGTTATCCA
2601 TTTTGAAGC AGCGCTTTTA CCGGACAAAT CAGCGGCGGC AAGGATACGG
2651 CATTAACATT AAAAGACAG GAATGGACGC GAATGGACGC TGGCTCAGG CACGGAATTA
2701 GCAATTTTAA ACCTTGACAA CGCCACCATT ACATCTCAAT CGCGCTATGG
60 2751 CACGATGCG CGAGGGGCGG AAACCGGCG TGCACAGAT GCGCGCGGCG
2801 CGCGTTCGCG CCGTTCCGCG CCGTCCCTAT TATCCGTAC ACGCGCAACT
2851 TCGCTACAGT CCGTCTGCAA CAGCTGAGC CAGCTGAGC AATTGACAGG
2901 TCAGGAGACA TTCCGCTTTA TGTGCGAAT TCTCGGTAC CCGAGGCA
2951 AATTGAAGCT GCGGAAAGT TCGGAAGCA CTTACACTTT GCGCGTCAAC
3001 AATACCGGCA ACGAACCTCG AAGCCTCGAA CAAATTGACG TAGTGAAGG
65 3051 AAAAGACCAAC AAACCGCTGT CCGGAAACCT TAATTTCACG TGTCAAAGC
3101 AACACGTGGA TGGCGGCGCG TGGCGTACCC AACTCATCCG CAAAGACGGC
3151 GAGTTCCGCC TGCATATACC GGTCAAGAAA CAAAGAGCTTT CCGACAAACT
3201 CGGCAAGGCA GAAGCAGAAA AACAGGCGGA AAAAGACAC GCGCAAGGCC
3251 TTGACGCGCT GATTGCGGCC GGGCGGATG CCGTCTGAAA GACAGAAAGC
70 3301 GTTGCGGAAC GCGCCGGGCA GCGAGCGGGG GAAATGTGTC GCATTATGCA

3351	GGCGGAGGAA	GAGAAAAAAC	GGGTCAGGC	GGATAAAGAC	ACCGCCTTGG
3401	CGAAACAGCG	CGAAGCGGAA	ACCCGCCGCG	CTACCACCGC	CTTCCCGCCG
3451	GCCCGCCGCG	CCCGCCGGGA	TTTCCGCGAA	CTGCAACCCC	AACCGCAGCC
3501	CCAACGCGAG	CGCGACCTGA	TCAGCCGTTA	TGCCAATAGC	GGTTTGAGTG
3551	AATTTTCCGC	CACGCTCAAC	AGCGTTTTCG	CCGTACAGGA	CGAATTAGAC
3601	CGCGTATTTC	CCGAAGACCG	CCGCAACGCC	GTITGGACAA	GCGGCATCCG
3651	GGACACCAAA	CACCTACCGT	CGCAAGATTT	CCGCGCTTAC	CGCCCAACAA
3701	CGGACCTGGC	CCAAATCGGT	ATCGAGAAAA	ACCTCGGACG	CGGGCGCGTC
3751	GGCATCTCGT	TTTCGACGAA	CCGCAACGAA	AGACCTCTCG	ACGACCGCAT
3801	CGCGAAGCTG	GCACGCGTTG	CCGACGCGCG	CGTTTTCGGG	CAATACGCGA
3851	TCGACAGGTT	CTACATCGGC	ATCAGCGCGG	GCGCGGGTTT	TAGCAGCGGC
3901	AGCCTTTTCG	ACGGCATCGG	AGGCAAAATC	GCGCGCGCGC	TGCTGCATTA
3951	CGGCATTTCG	GCACGATACC	GCGCGGTTT	GCGCGGATTC	GGCATCGAAC
4001	CGCACATCGG	CGCAACGCGC	TATTTCTGCC	AAAAAGCGGA	TTACCGCTAC
4051	GAAACAGTCA	ATATCGCCAC	CCCGCGCCTT	GCAATTCACAC	GCTACCGCGC
4101	GGGCATTAA	GAGATATTAT	CATTCAAAAC	GGCGCAACAC	ATTTCTCATCA
4151	CGCCTTATTT	GAGCCTGTCC	TATACGATCG	CGCCTTCGGG	CAAAAGTCCGA
4201	ACACGCGTCA	ATACCGCGCT	ATTGGCTCAG	GATTTTCGCA	AAACCCGCGAC
4251	TGCGGAATGG	GGCGTAAACG	CGGAAATCAA	AGGTTTTCACG	CTGTCCCTCC
4301	ACGCTGCGCG	CGCCAAAGCG	CGCAACTCGG	AAGCGCAACA	CAGCGCGGGG
4351	ATCAAAATTAG	GCTACCGCTG	GTA		

This corresponds to the amino acid sequence <SEQ ID 650; ORF1-1>:

1	MKTTDKRTTE	THRKAPKTGR	IRFSPAYLAI	CLSPGILFOA	WAGHTYVGIN
51	YQYYRDFAE	KGKFAVGA	KEVYNGKGL	VGKSMTKAPM	IDFSVVRNG
101	VAAVLVDQYI	VSAVHNGY	NVDFGAEGRN	FDQHRFTYKI	VKRNRYKAGT
151	KGHPYGGDYH	MFLRHKFTVD	AEPVEMTSM	DGRKYIDQNN	YFDRVRIGAG
201	RQYWRSEDE	PNNRSESYH	ASAYSWLVG	NTFAQNGSGG	GTVNLGSEKI
251	KHSPYGFLEP	GGSPGDSGSP	MFYIDAQKQ	WLINGVLQGT	NPYIGKSNFG
301	QLVRKWFYD	EIFAGDTHSV	FYEPRQNGKY	SFNDNNNGTG	KINAKHEHNS
351	LPNRLKTRTV	QLPNVSLSET	AREPVMHAAG	GVNSYRPRLN	NGENISFIDE
401	KGKELILTSN	INQAGAGLYF	QGDFTVSPEN	NETWOGAVGH	ISEDSTVTWK
451	VNGVANDRLS	KIGKGLTHVQ	AKGENQGSIS	VGDGTVIDLQ	QADDKGRKQA
501	FSEIGLVSGR	GTGVLNADNQ	FNFDKLYFGF	RGRRLDLNGH	SLSFHRIQNT
551	DEGAMLVNHN	QKEISYVFTT	ENKDEATTGN	NNSLDSKKEI	AYNGWFGKED
601	TKTKNGRLAL	VYQFAAEEDT	LLLSGSGTALN	GNITGTFEHA	PFSGRTEFHA
651	YNHLNDHWSQ	QKGLPRGEIV	WDMNDWNRFT	KAEKFIKGG	QAVYSKNVAK
701	VKGWHLNSNH	AQAVFGVAPH	QSHTICTRSD	WTGLTNCEVK	TITDDKVIAS
751	LTKTDISGNY	DLADHAHLNL	TGLATLNGNL	SANGDTRYTV	SHNATQNGNL
801	SLVGNAGATF	NQATLNGNTS	ASGNASFNLS	DHAVONGSLT	LSGNKAGNVS
851	HSALNGNVSL	ADKAVFHES	SRPTGQISGG	KDTALHLKDS	EWTLPSGTGL
901	GNLNLNATII	TLSAYRHDA	AGAQTGSATD	APRRSRRSR	RSLLSVFTPT
951	SVESRNTLT	VNGKLANGQT	FRFMSLFY	RSDDLKLAES	SEGTYTLAVN
1001	NTGNEPASLE	QLTVVEGKDN	KPLSENINFT	LQNEHVDAGA	WRYQLIRKDG
1051	EFRLHNPVKE	QELSDKLGA	EAKQKAEKDN	AQSLDALAA	GRDAVEKTES
1101	VAFEPARQAG	ENVMGMAZEE	EKKRVQADKD	TALAKORAE	TRPATTAFPR
1151	ARRARRDLFQ	LQFGQFGQFG	RDLISRYANS	GLSEFSATLN	SVFAVDEDEL
1201	RVFAEDRRNA	VMTSGIRDTK	HYRSQDFRAY	RQOTDLRQIG	MOKNLGSGRV
1251	GLFSGHNRTE	VTDFGIGENS	RLAIDAVGFG	QYGIIDRFYIG	ISAGCAQFSG
1301	SLSDGIGGKI	RRRVLYHGIO	ARYRAGPGGF	GIEPHIGZRY	YVDASGKVR
1351	ENVNIATPGL	AFNRYRAGIK	ADYSKCPAQH	ISITPYLSLS	YVDASGKVR
1401	TRVNTAVLAQ	DFGKTRSAEW	GVNAEIKGFT	LSLHAAAKG	PQLEAQSBSAG
1451	IKLYRW*				

Computer analysis of these sequences gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

- 55 ORF1 shows 57.8% identity over a 1456aa overlap with an ORF (ORF1a) from strain A of *N. meningitidis*:

		10	20	30	40	50	60
orf1.pep		MKTTDKRTTETHRKAPKTGRIRFEXAAYLAICLSFGILFOAWAGHTYVGINVQYYRDFAE					
60 orf1a		MKTTDKRTTETHRKAPKTGRIRFSEPAYLAICLSFGILFOAWAGHTYVGINVQYYRDFAE					
		10	20	30	40	50	60
orf1.pep		KGKFAVGAKEIVYNGKGLVGSMTKAPMIDFSVVRNGVAALVGVQYIVSAHNGY					

10

15

20

26

30

35

40

45

50

55

60

65

70

490 500 510 520 530 540
orfl.pep GDTRYTVSHNATQNGNXSLVXNAQATFNCATLNGTSSASGNASFNLSDHAVONGSITLSC

10

15

20

22

30

35

40

45

50

55

65

70 The complete length ORF1a nucleotide sequence <SEO ID 651> is:

1 ATGAAACAA CCGACAAACG GACAACCGAA ACACACCGCA AAGCCCGAA
 51 AACCGGCGCG ATCCGCTTCT CGCGTGCTTA CTTAGCCATA TGCCTGCGT
 101 TCGGCATTCT TCCCAAGCT TGGGCGGGAC ACACATATTT CGGCATCAAC
 151 TACCATACT ATCCGCGACT TGGCGAAAT AAGGCGAAT TTGCGATCGG
 201 GCGCAAGAT ATTAGGNTT ACACAAAAA AGGGGAGTTG GTGCGCAAT
 251 CAATGACAAA AGCCCCGATG ATTGATTTTT CTGTGGTGTC GCGTAAOCCG
 301 TGGCGGCAT TGGTGGGGGA TCAATATATT GTGAGCGTGG CACATAACGG
 351 CGGCTATAC AACCTTGATT TTGGTGCGGA AGGAGNGAAT CCGCATCAGC
 401 ACGGTTTTCT TACCAAAAT GTGAAAGAA ATAATTATA GCTTGCAANT
 451 TCGACCGCTT ACACCGGCGA TTAACATATG CCGGTTTGG ATAAATTTGT
 501 CACAGATGCA GAACCTGTCT AAATGACGAG TGACATAGGG GGGATACCT
 551 ATTCCGATAA AGAAAAATAT CGCGAGCGTG TCCGATCGG CTCAGGACAC
 601 CACTATTGGC GTTATGATGA TGACAAACAC GGCATTTAT CTTACTCCGG
 651 CGCATGGTTA ATTGGCGGCA ATACACATAT CGAGGGTTGG GGAATAAATG
 701 GGTANTTAG TTGAGCGGCG GATGTGCGCC ATGCCAACGA CTATGGCCCT
 751 ATGCGGATTG CAGGTGCGGC AGGCGACAGC GGTTCGCCAA TGTTTATTTA
 801 TGACAAAAA AACAATAAAT GGCTGCTCAA CGGAGTTTTA CAAACCGGCT
 851 ACCCTTATTC CGGCAGGGAA AAGCGTTTCC AGTGATACG CAAAGATTGG
 901 TTCTACGATG ACATTTACAG AGGCGATACA CATACCGTCT NTTTTGAACC
 951 GCGCAGTAAC GGACATTTTT CTTTACATC ACACAAACAC GGTAGGGGTA
 1001 CGGTACCGGA AACCACGAA AAGGNTTCCA ATCCAAAGCT TAAGTACAGT
 1051 ACAGTCGCGA TGTTTGAGGA ATCTTTGAT GAACCTGATA AAGAACCTAG
 1101 TTACGCGGCA GGGGTTGTTA ATCACTACGC TCCAAAGTTA AAGACCGG
 1151 AAACCTTTCT TTTTATCGAT TACGGCAACG GCAACCTCAT CTATCAAC
 1201 AACATCAACC AAGGCGCGGG CGGTTTGTAT TTTGAGGTG ATTTTACGGT
 1251 CTGCGCTGAA AACAAAGAAA CGTGCGAAGC CGCGGGCGTT CATATCAGTG
 1301 AAGACAGTAC GCTTACTTGG AAGTAAACG CGGTGCGAAA CGACCGCGTG
 1351 TCCAAATACG CAAAGGCGAC GCTGACAGTT CAAGCCAAAG GGGAAACCA
 1401 AGGCTCGATC AGCGTGGGCG ACGGTACAGT CATTTTGTAT CAGCAGGCGAG
 1451 ACGATAAAGC CAAAAACAA GCCTTTAGTG AAATCGGCTT GNTCAGCGGC
 1501 AGGGGTACGG TGCAACTGAA TGGCGATAAT CAGTTCAACC CGACAAACT
 1551 CTATTTCCGC TTTGCGGGCG GACGTTTTGA TTTAAACGGG CATTGCTTT
 1601 CGTTCACACG TATTTAAAT ACCGATGAAG GGGCGATGAT TGNCAATCAT
 1651 AATGCAACGA CACATCACT CCGTACCATG ACAGGGAATG AAGTATGAT
 1701 ACACACAGCT GCTAAGATA TCAATAGATT TAATTTACAG TAACAAATTTG
 1751 CTTACAACGG TTGTTTGGC GAGAAAGATA GACCAAAAC GAACGGGCGG
 1801 CTCACACTTG TTTACAGCC CGCGCGAGAA GACCGCACCC NGCTGCTTTC
 1851 CGCGGGAAACA AATTTAAAG GCAACATCAC GCAACAAAC GGCAAACTGT
 1901 TTTTCAGCGG CAGACCGACA CGCGACGCT ACATCATTT AGGAAGCGGG
 1951 TGGTCAAAAA TGGAGGTTAT CCCACAAGGA GAATCGTGT GGGACAACGA
 2001 CTGATCAACG CGCACGTTTA AAGCGGAAAA TTTCCATATT CAGGCGCGGG
 2051 AGGCGGTGAT TTCCGCAAT GTTGCCAAAG TGGAGGCGCA TTGNCATTTG
 2101 AGCAATCAGC CCGAAGCAGT TTTTGTGTCT GCACCGGCATC AAGCGCATAC
 2151 AATCTGTACA GCTTCGAGT GGACNGGTCT GACAAATGT GTGCAANAAA
 2201 NCATTACCGA CGATAAAGTG ATTGCTTCAT TGACTAAGAC NGACNTNAGC
 2251 GGCATATGNA GNTNNNCCTA ATTTGAGCTT TNAAGAACT CNGGCGNCTG
 2301 NNCATTAAN GCGACTCTTA GTGCAATGT CGATACAGT TATACAGTCA
 2351 GCCACACAGC CACCAAAAC GGCACACTTA GCGCTGTGGG CAATGCCCAA
 2401 GCACACTTTA ATCAAGCCAC ATTAAGGCGC AACNACTCG NTTGCGGCAA
 2451 TGCTTCATTT AATCTAAGCA ACAGGCGCGC ACACAAACGG AGTCTGACGC
 2501 TTTCCGACAA CGCTAAGGCA AAGCTAAGCC ATTCCGCACT CAACGGCAAT
 2551 GTCTCCCTAG CGATAAGGC AGTATTCCAT TTTGAAACCA GCGCTTTTAC
 2601 CGGACAACTC AGGCGGACGA AGGANACAGC ATTACCTTA AAGACAGCG
 2651 AATGGAAGCT GCGCTCAGGC ACGGAATTAG GCAATTTAAA CTTTGCAAC
 2701 GCCACATTA CACTCAATTC CGCCTATGCG CACGATGCTG CAGGCGGCGA
 2751 AACCGGCGAG GTCTCAGACA CGCGCGCGCG CGGTTGCGCG GCTTCCCTAT
 2801 TATCCGTTAC ACCGCGCACT TCGGTAGAAT CCGGTTTCAA CACGCTGACG
 2851 GTAAACGGCA AATTGAACNG TCAAGGAACA TTCCGCTTTA TGTGCGAAT
 2901 CTTGCGCTAC CGAAGCGACA AATTGAAGCT GCGGGAAGGT TCCGAAGNA
 2951 CTTGACACTT GCGACTCAGC AATACCGCA AGACACCGT ACCTGATGAT
 3001 CAATTGACGG TAGTGGAGG GAGACACAC AAACCGCTGT CCGAANAACCT
 3051 TAATTTCAAC CTGCAAAAGC AACACGTGCA TGCAGCGCGC TGGGCTTACC
 3101 AACTCATCCG CAAAGACCGC GAGTTCCGCG TGACATAAAT GGTCAAAGAA
 3151 CAAGAGCTTT CGGCAAACT CGGCAAGGCA GAAGCCAAAA AACAGCGGGA
 3201 AAGAGCAACG GCGCAAGGCC TTGACGCGCT GATTGCGGCG GGGCGGATG
 3251 CGCGCGAAAA GACAGAAAGC GTTGCGGAC CGGCGCGGCG GCGAGCGGG
 3301 GAAATGTGCG GATTTATGCA GGGCGAGGAA GAGAAJAAAC GGGTGCAGCG
 3351 GGATAAAGAC AGCGCNTTGG GAAACAGCG CGGAGCGGAA ACCGCGCGG
 3401 NTACCAACCG CTTCCCGCGC GCGCGGCGCG CCGCGCGGGA TTTGCGGCAA
 3451 CGCAGAGCCC AACCGCAACC TCAACCCCAA CGCAGCGCGC ACCTGATNAG
 3501 CGCTTATGCG AATAGCGGTT TGAATGAATT TTCCGCGCAG CTCACAGCGG
 3551 TTTTCCGCGT ACAGGACGAA TTGGACCGTG TTTGCGCGTA AGACCGCGCG

3601 AACGCNGTTT GGACAAGCNG CATCCGNGAC ACCAAACACT ACCGTTGCGA
 3651 AGATTTCCGC GCCTACCGCC AACAAACCGA CCTGCGCCAA ATCGGTATGC
 3701 AGAAAAACCT CGCGACGGGG CGCGTCGGCA TCTGTTTTC GACAACCGG
 3751 ACCGAAACCA NCTTCGACGA CGGCATCGGC AACTCGGCAC GGCTTGCCCA
 3801 CGCGCCCGTT TTCGGGCAAT ACGGCATCGG CAGGTTCCAG ATCGGCATCA
 3851 GCACGGGCGC GGGTTTATAG ACGGCATCCT TMTCAGACGG CATCGGAGCG
 3901 AAAATCCGCG CGCGCGTGTG GCATTACGGC ATTACGACAC GATACCGCGC
 3951 CGGTTTCGGC GGATTTCGGCA TCGAACCGTA CATCGGCGCA ACGCGCTATT
 4001 TCGTCACAAA AGCGGATATG CGCTACGAAA ACCTCAATAT CGCCACCGCC
 4051 GCTCTTCGCT TCAACCGGTA CCGGCGGGCG ATTAAAGCAG ATTTATTCAT
 4101 CAAACCGCGC CAACACATNT CCATCAACNC TTATTTNAGC CTGTCCTATA
 4151 CCGATCGCGC TTCGGGCAAA CTGCGAACAC CGCTCAATAC CGCNGTATTG
 4201 GCTCAGGATT TCGCGAAAAA CCGCATGTGG GAATGGGGCG TAAACGCCGA
 4251 AATCAAGAGT TTCACGCTGT CMTCCACGC TGCGCGCGCC AAAGNCCGCG
 4301 AACTGGAAGC GCAACACAGC GCGGGCATCA AATTAGGCTA CCGCTGGTAA

This encodes a protein having amino acid sequence <SEQ ID 652>:

1 MKTTDKRTTE THRKAAPTGR IRFSPAYLAI CLSPGILPOA WAGHTYFGIN
 51 YQYRDFAEK KGFVAGAKD IEVYNKKGEL VGKSMTKAPM IDFSVVSRRNG
 101 VAALVGDQYI VSVAHNGGYN NVDFGAEKGN PQQHRFSYQI VKRNNYKPDN
 151 SHPYNGDXHM PRLHKFVTD AEPVEMTSDMR GNTYSDEKYE PERVRIGSGH
 201 HWRYDDDKH GDLISYGAWL ICGMTHMQGW GNGVGSLSG DVRHANDIGP
 251 MPLAGAGDS GSPEYVDTK NKKWLLANGYL QGVPYSGRE NGFQLIRKQ
 301 FYDDIYRGDT HTVXFEPBSM GHFSPTSNWN GGTVTVTNE KVENPKLVQ
 351 TVRLFDIESLN ETQKEPVYAA GGVMQYRPL NNGENLSFD YGNGKLILSN
 401 NINQAGAGLY FEGDFTVSP NNETWQAGV HISEDSTVW KVNQVANDRL
 451 SKI GKGTLHV QAKGENQSI SVGDGTVID QOADDKGGKQ AFSEIGLXSG
 501 RGTVLNADN QFNPKLYFG FRGRLDLNG HSLFHRION TDEGAMIXXH
 551 NATTTSTVTI TGNESITQPS GKNINRLNYS KELAYNGWFG EKDPTKTNGR
 601 LNLVYQPAAE DRTXLLSGGT NLNGNIQTQ NKLFYSGRPT PHAYNHLSSG
 651 WSKMEGIPQG ELVDNDWIX RTFKAENFHI QGGQAVISR NAKVEGDHXL
 701 SNHAQAVFG APHQSHICT RSDWTGLTNC VEXXITDDKV IASLTXTDXS
 751 GXVXLXXXXX XXILGXKXIL GNLSANGDTR YTVSHNATCN GNLSLVGNAQ
 801 ATFNQATLNG NXSXGNASF NLSNNAQNG SILTSDNKA NVSHSALANG
 851 VSLADKAVFH FENSRETGQL SSKXTALHL KSEWTLPEG TELGNLNDN
 901 ATTILNAYR HDAQAQNGX VSDTPRRRS RLLSVTPEP VESRFTVL
 951 VNGKLNKQGT FRMSELFYV RSDKLKLES SEGTYPLAVN NTCNEPVSLD
 1001 QLTVVEGKDN KPLSENLMFT LQNEHVDAGA WRYQLIRKDG EFRLNHPVKE
 1051 QELSDKLKGA EAKKQAEKDN AQSLOALIA GRDAEKTES VAEFAPXAGG
 1101 ENVGIMQAE EKRRVQADKO SALAKOREAE TRPXTTAFPR ARXARRDLPO
 1151 PQPQPQPQPQ PQDRLXSYA NSGLSEFSAT LNSVFAVDGE LDRVFAEDRR
 1201 NAVWTSXIRX TKHYRSQDPR AYRQQTDLRG IGMQKNLGSQ RVGILFESHNR
 1251 TENXFDGIGX NSARLAHGV FGQYIGIRFD IGISTGAGFS SGXLSDGIGG
 1301 KIRRRVLHYG IQARYRAGFG GFGIEPIYA TRYFVQKADY RYENVNIAFP
 1351 GLAFNRYRAG IKADYSFKPA QHSITPYXS LSYTDAASGK VRTVRNPAVL
 1401 AQDPGKTRSA EWGVNAEIKG FTLSXHAATA KGPOLAEQHS AGIKLGYRW*

A transmembrane region is underlined.

ORF1-1 shows 86.3% identity over a 1462aa overlap with ORF1a:

10 20 30 40 50 60
 orf1a.pep MKTTDKRTTETHRKAAPTGRIRFSPAYLAICLSFGILPOAWAGHTYFGINQYRDFAEK
 50 orf1-1 MKTTDKRTTETHRKAAPTGRIRFSPAYLAICLSFGILPOAWAGHTYFGINQYRDFAEK
 10 20 30 40 50 60
 70 80 90 100 110 120
 orf1a.pep KGFVAGAKDIEVYNKKGELVGKSMTKAPMIDFSVVSRRNGVAALVGDQYIVSVAHNGGYN
 55 orf1-1 KGFVAGAKDIEVYNKKGELVGKSMTKAPMIDFSVVSRRNGVAALVGDQYIVSVAHNGGYN
 70 80 90 100 110 120
 130 140 150 160 170 179
 orf1a.pep NVDFGAEKGNPQQRFSYQIVKRRNNYKPDNS-HPYNGDXHMPRLHKFVTD AEPVEMTSDMR
 60 orf1-1 NVDFGAEKGNPQQRFTYKIVKRRNNYKAGTKHPYGGDYHMPRLHKFVTD AEPVEMTSDMR
 130 140 150 160 170 180

		180	190	200	210	220	230
	orf1a.pep	RGNTYSDEKQYPERVIRIGSGHHYWRYYDDDKHGD--SYSGA---WLIIGGNTIMSGWGN					
5	orf1-1	DGRKYIDONNYPDRVIRIGAGROYRSDDEENNRESSYHIASAYSWLVGNTFAGNGSGG					
		190	200	210	220	230	240
	orf1a.pep	240	250	260	270	280	290
10	orf1-1	GVXSLSGD--VRHNDYGPMPFAGAAAGDSGSPMFYDKTNNKWLINGVLQGYPPSGRENG					
		250	260	270	280	290	
	orf1a.pep	300	310	320	330	340	350
15	orf1-1	FOLIRKDFYDDIYRGDTHVXFEPRNGHFSFTSNNGTGTVTETNEKVSNP--KLKVQT					
		310	320	330	340	350	
	orf1a.pep	360	370	380	390	400	410
20	orf1-1	VRLFDESINMETKEPVY--AAGGVNQYRPLNNGENLSFYDYGNGKLLSNNINQAGGGLY					
		360	370	380	390	400	410
	orf1a.pep	420	430	440	450	460	470
25	orf1-1	FEGDFTVSPENNETWQAGVHIHSEDSTVTWKVNGVANDRLSKIGKGLTHVQAKGENQSGI					
		420	430	440	450	460	470
	orf1a.pep	480	490	500	510	520	530
30	orf1-1	SVGDGTVLDDQADDKGGKQAFSEIGLXSGRGTVQLNADNQFNPDKLYFGFRGGRRLDLNG					
		480	490	500	510	520	530
	orf1a.pep	540	550	560	570	580	590
35	orf1-1	HSLSFHRIQNTDEGAMIXXHNATTTSTVTITGNE:ITQPSGKNINRLNYSKEIAYNGWFG					
		540	550	560	570	580	590
	orf1a.pep	600	610	620	630	640	650
40	orf1-1	EKDTTKTNGRLNLVYQPAEDRTKLLSGGTNLNGNITQTNGKLFFSGRPTPHAYNHLGSG					
		600	610	620	630	640	650
	orf1a.pep	660	670	680	690	700	710
45	orf1-1	WSKMEGIPQGEIVWDNQWIXRTTFAENFHIQGGQAVISRNVAKVGGXHLNHAQAVFGV					
		660	670	680	690	700	710
	orf1a.pep	720	730	740	750	760	770
50	orf1-1	APHQSHTICTRSDWTGLTNCVEXXITDDKVIASLTKTDXSGXVLXXXXXXLXGKAXLK					
		720	730	740	750	760	770
	orf1a.pep	780	790	800	810	820	830
55	orf1-1	GNLSANGDTRYTVSHNATQGNLSLVGNAQATFNQATINGNXSXSGNASFNLSNNAQNG					
		780	790	800	810	820	830
	orf1a.pep	840	850	860	870	880	890
60	orf1-1	SITLSGNNAKAVNHSALNGVSLADKAVFHFE:SRFTGQISGGKDTALHLKDEWTLPSG					
		840	850	860	870	880	890

		900	910	920	930	940
	orf1a.pep	TELGNLNLNDNATI TLNSAYRHDAAAGQTGVSDTFRRRSRSS---LLSVTPPTTSVESRFN				
5	orf1-1	TELGNLNLNDNATI TLNSAYRHDAAAGQTGSATDAPRRRRSRSSRLSVTPPTTSVESRFN				
		900	910	920	930	940
	orf1a.pep	950	960	970	980	990
10	orf1-1	950	960	970	980	990
		960	970	980	990	1000
	orf1a.pep	1010	1020	1030	1040	1050
15	orf1-1	1010	1020	1030	1040	1050
		1020	1030	1040	1050	1060
	orf1a.pep	1070	1080	1090	1100	1110
20	orf1-1	1070	1080	1090	1100	1110
		1080	1090	1100	1110	1120
	orf1a.pep	1130	1140	1150	1160	1170
25	orf1-1	1130	1140	1150	1160	1170
		1140	1150	1160	1170	1180
30	orf1a.pep	1190	1200	1210	1220	1230
	orf1-1	1190	1200	1210	1220	1230
		1200	1210	1220	1230	1240
35	orf1a.pep	1250	1260	1270	1280	1290
	orf1-1	1250	1260	1270	1280	1290
		1260	1270	1280	1290	1300
40	orf1a.pep	1310	1320	1330	1340	1350
	orf1-1	1310	1320	1330	1340	1350
		1320	1330	1340	1350	1360
45	orf1a.pep	1370	1380	1390	1400	1410
	orf1-1	1370	1380	1390	1400	1410
		1380	1390	1400	1410	1420
50	orf1a.pep	1430	1440	1450		
	orf1-1	1430	1440	1450		
55		1440	1450			

Homology with adhesion and penetration protein hap precursor of *H. influenzae* (accession number P45387)

Amino acids 23-423 of ORF1 show 59% aa identity with hap protein in 450aa overlap:

orf1	23	FXAAYLAICLSFGILPQAWAGHTYFGINYYQYRDFAEKNGKFVAVGAKDIEVYNNKKGLVG	82
		F +L C+S GI QAWAGHTYFGI+YQYRDFAEKNGKF VSAK+IEVYNNK+G+LVG	
65	hap	FRNLNLTACVSLGASQAWAGHTYFGIDYQYRDFAEKNGKFVYVARNIEVYNNKKGLVG	65
orf1	83	KSMTKAPMIDFVSVSRNGVAALVGQYIVSVAHNGGYNVDFGAGKXNIXQXRTYKTV	142
		SMTKAPMIDFVSVSRNGVAALVG QYIVSVAHNGGYN+VDFGAEK N DQ R TY+IV	
hap	66	TSMTKAPMIDFVSVSRNGVAALVGQYIVSVAHNGGYNVDFGAEGRN-PQQRHRTYQIV	124

5	orf1	143	KRNNYKAGTKGHPYGGDYHMPRLHKXVTDAPVEMTSYMDGRKYIDQNNYPDRVRIGAGR	202
	hap	125	KRNNYQAWERKHYPYGGDYHMPRLHKFVTEAEFVGMITNMDGKVYADRENYPERVIRIGSSGR	184
	orf1	203	QYWRSDDEFPNNRESSYHIA-----	222
	hap	185	QYWRD+DE N SSY++	244
10	orf1	223	-----SGSPMFYIDAKQKQWLINGVLQTNPNYIGKSNQFOLVRKDMFYDEIFAGDTHSVF	277
	hap	225	SGSPMFYIDA+K+WLIN VLQGT+P+ G+ NGFQL+R++WFY+E+ A DT SVF	304
15	orf1	278	--YEPQNGKYSFNDNNNGTGKIN-AKHEHNSLPNRLKTRTVOLFNVSLSETAREPVPYHA	334
	hap	305	Y P NG YSF +N+GTGK+ + + + TV+LFN SL++TA+E V A	363
20	orf1	335	AGGVNSYRPLNNGENISFTIDEGKGLILTSNINQAGAGGLYFQGDFTV-SPENNETWQGA	393
	hap	364	A G N Y+PR+ G+NI D+GKG L + +NINQAGAGGLYF+G+F V +NN TWQGA	423
	orf1	394	GVHISEDSTVTWVKNVANDRLSKIGKCTL 423	
	hap	424	GV I +D+TV WKV+ NDRLSKIG CTL	453
25	Amino acids 715-1011 of ORF1 show 50% aa identity with hap protein in 258aa overlap:			
	Orf1	41	DTRYTVSHNATQ-NGNXSLVXNAQATFNO-ATLNGNTSASGNASFNLSHRAVQNGSLTS	98
	hap	733	DT+ S TQ NG+ +L NA + A LNGN + ++ F LS++A Q G++ LS L N	792
30	orf1	99	GNAKANVSHSALNGNVSLADKAVHPFESSRFTGQISGGKDTALHLKDSWTLPSSXELGN	158
	hap	793	+A A V+++ LNGNV L D A F ++S F QI G KDT + L+++ WT+PS L N	852
35	orf1	159	LNDNATITLNSAYRHDAAGAQTSATDAPXXXXXXXKLLXVTPPTSVESRNTLTVN	218
	hap	853	L L+N+T+TLNSAY + S+ +AP L T PTS E RNTLTVN	899
40	orf1	219	GKLNQGGTFRFMSELFGYRSOKLKLAESEGGTYTLAVNNTGNEPASLEQLTVVEGKDNKP	278
	hap	900	GKL+GGGTFF+ F S LFGY+SDKLKL+ +EG YTL+V NTG EP +LEQLT++E DNKP	959
	orf1	279	LSENLNFQLNEHVDAGA 296	
	hap	960	LS+ L FTL+N+HVDAGA	977
45	Amino acids 1192-1450 of ORF1 show 41% aa identity with hap protein in 259aa overlap:			
	Orf1	1	LDRVFAEDRRNAVWTSIGRIDTKHYRSODFRAYRQQTDLAIGMOKNLGSGRVGILFHSNR	60
	hap	1135	LDR+P + ++AVWT+ +D + Y S PRAY+Q+T+LRQIG+QK L +GR+G +PSH+	1194
50	orf1	61	TENTFDDGIGNSARLAHGAVFQYQIDRFYXXXXXXXKXKXRRRLVHYG	120
	hap	1195	+N+TFD+ + N A L + F QY K R+ +YG	1254
55	orf1	121	IQARYRAGFGGPGIEPHIGATRYFVQKADRYRNVNIATPGLAFNRYRAGIKADYSFKPA	180
	hap	1255	+A Y+ G GT+P+G RYF+++ +Y+ E V + TP LAFNRY AGI+ DY+P P	1314
60	orf1	181	QHISITPYLSLSYTDASGKVRTRVNTAVLAQDPFGKTRSAEAGVNAEIKGFTLSLHAAA	240
	hap	1315	+IS+ PY ++Y D + +V+ VN VL Q FG+ E G+ AEI F +S + +	1374
	orf1	241	KGPQLEAQSAGIKLGYRW 259	
	hap	1375	+G QL Q + G+KLGYSW	1393
65				

Homology with a predicted ORF from *N.gonorrhoeae*

The blocks of ORF1 show 83.5%, 88.3%, and 97.7% identities in 467, 298, and 259 aa overlap, respectively with a predicted ORF (ORF1ng) from *N.gonorrhoeae*:

5	orf1. pep	MKTDKRTTETHRKAPKTGRIRFXAAYLAICLSFGILPQAWAGHTYFGINQYQYRDFARN	60
	orf1ng	MKTDKRTTETHRKAPKTGRIRFSPAYLAICLSFGILPQARAGHTYFGINQYQYRDFARN	60
10	orf1. pep	KGKFAVGAKDIEVYNNKGELVGKSMTKAFMIDFSVSRNGVAALVGVOYIVSVAHNGGYN	120
	orf1ng	KGKFAVGAKDIEVYNNKGELVGKSMTKAFMIDFSVSRNGVAALVGDOYIVSVAHNGGYN	120
15	orf1. pep	NVDFGAEGXNIXDQXRTYKIVKRNNYKAGTKGHFYGGDYHMPRLHKXVTDAEPVEMTSY	180
	orf1ng	NVDFGAEGSN-PDQHRFSYQIVKRNNYKAGTNGHFYGGDYHMPRLHKFVTDAEPVEMTSY	179
20	orf1. pep	MDGRKYIDQNNYPDRVRIGAGRQYWRSEDEPNNRESSYHIAS-----	223
	orf1ng	MDGWKADLNKYDPRVRIGAGRQYWRSEDEPNNRESSYHIASYSWLVGNGTFAQNGSG	239
25	orf1. pep	-----GSPMFYDAQKQKWLINGVLQTNPIYKSGNSG	255
	orf1ng	GCTVNLGSEKIKHSPYGFLLPTGSGSGSPMFYDAQKQKWLINGVLQTNPIYKSGNSG	289
30	orf1. pep	<u>EQLVRKDWFYDEIFAGDTHSVFYEPQNGKYSFNDDNNGTGKINAKHEHNSLNRILKTRT</u>	315
	orf1ng	<u>EQLVRKDWFYDEIFAGDTHSVFYEPHNGKYSFNDDNNGAGKIDAKHKHYSLEPYRLKTRT</u>	359
35	orf1. pep	VQLFNVSLETAREFVYHAAGGVNSYRPLNNGENISFIDEGKGELILTSNINQAGGLY	375
	orf1ng	VQLFNVSLETAREFVYHAAGGVNSYRPLNNGENISFIDRKGKELILTSNINQAGGLY	375
40	orf1. pep	FGQDFVSPENNETWQAGVHISDSTVTWKVNGVANDRLSKIGKT	422
	orf1ng	FEQNFVSPKNNETWQAGVHISDSTVTWKVNGVANDRLSKIGKFTLLVQAKGENQGSV	479
45	orf1. pep	//	
	orf1ng	DKVTASLTKTDISGNVDLADHAHLNLTGLA	744
50	orf1. pep	FGVAPHQSHTICTRSDWTGLTSCTEKITDDKVIASLSKTDVRGNVSLADHAHLNLTGLA	774
	orf1ng	FGVAPHQSHTICTRSDWTGLTSCTEKITDDKVIASLSKTDVRGNVSLADHAHLNLTGLA	774
55	orf1. pep	TINGNLSANGDTR-YTVSHNATQNGNKSILVXNAQATFNQATLNGTNSAGSNASFNLSGHA	803
	orf1ng	TFNGNL-VQAEFTRTIRLANATQNGNLSLVGNAQATFNQATLNGTNSAGSNASFNLSGNA	833
60	orf1. pep	VQNGSLTSLGNKANKVSHSALNGVSLADKAVHFHFNSSRFTGQISGGKDTALHLKDSWT	863
	orf1ng	VQNGSLTSLDNKANKVSHSALNGVSLADKAVHFHFNSSRFTGQISGGKDTALHLKDSWT	893
65	orf1. pep	LPSGXELGNLNDNATITLNSAYRHAAGAQTGSATDAPRRRRRRSRLSLXVTPPTSVE	923
	orf1ng	LPSGTELGNLNDNATITLNSAYRHAAGAQTGSADAPRRRRRRS---LLSVTPPTSVE	950
70	orf1. pep	SRPFWTLVNGKLNQGTFRFMSELFYGRSDKLLAESSEGYTLAVNNTGNEPASLEQLT	983
	orf1ng	SRPFWTLVNGKLNQGTFRFMSELFYGRSGKLLAESSEGYTLAVNNTGNEPVSLEQLT	1010
75	orf1. pep	VVEGKDNKPLSENINFTLQNEHVDAGW	1011
	orf1ng	VVEGKNTPLSENINFTLQNEHVDAGAWRYQLIRKDGFEFLHNPVKEQELSDKLGKAGET	1070
80	orf1. pep	//	
	orf1ng	LDRVFAEDRRNAVWTSIGIRDTKHYSQDFR	1211
85	orf1. pep	PQRDLISRYANGLSEFSATLNSVFAVQDELDRVFAEDRRNAVWTSIGIRDTKHYSQDFR	1239
	orf1ng	PQRDLISRYANGLSEFSATLNSVFAVQDELDRVFAEDRRNAVWTSIGIRDTKHYSQDFR	1239
90	orf1. pep	AYRQOTDLRQIGMQKNLGSGRVGILFSHNRTENTFDDGIGNSARLAHGAFFQYQIGDRFY	1271
	orf1ng	AYRQOTDLRQIGMQKNLGSGRVGILFSHNRTENTFDDGIGNSARLAHGAFFQYQIGDRFD	1299

15	1	ATGCAAAACAA	CGCAGCAATCG	GACACACACG	ACACAACCGA	AAGCCOCTAA
	51	AGGCGGCGCG	ATCCGCTCTT	CGCCCGCTTA	CTTAGGACAT	TGCGTCTCGT
	101	TGCGGATCTT	CGCCCAAGCC	CGCGGGAAAT	ACACATTTAT	CGGCATCAAC
	151	TACCAATTAAT	ATTCGGCAAT	TGCGCGCAAT	AAGAGGCAGT	TTCAGCTCGG
	201	CGGCAAAAGT	ATTAGGTTAT	ACACAACAAA	AGGGGATGTT	TGCGGCARAT
	251	GCGACGACAA	AGGCCCGATG	ATTGATTTAT	TTGTGTATCT	GGCTAAACGG
	301	TGGCGGCGAT	TGGCGGGCGA	TCAATTTAT	GTGACGGTAT	CGCATTAACGG
	351	CGGCTATPAC	ATTGTTGATT	TGCTGTGCGA	GGGAGCAAGT	CCGATCAAGC
	401	ACCGCTTTTC	TTACGAAATT	GTGAAAGAA	ATAATTATRA	AGCGAGGACT
	451	AGCGGCGCTC	TTATGCGGAT	CGATTATCAT	ATTCGCGGCT	TGCGCAAAAT
	501	TGTGACGAT	TTGAATCTCT	TGCGAGTGAT	TGCGGAGGAT	GATGGTGATC
	551	ATATCGATCT	TTTCTATAAT	TACCTGATCT	TGCTTCGAGT	CGGAGAGAGT
	601	AGACATATAT	GGCGTTCTGA	TGAGACAGAA	CCCATATPAC	CGGAAAGTTC
	651	ATCATATATT	CGACGCGCAT	ATTCTTGCTT	CGTCGCTGGC	ATACCTTTTG
	701	CACAAATAGG	ATCAGGTGGT	GGCATCTGAT	ACTTAGGTAG	CGAAARAAAT
	751	AAACATGACC	ATCATGGTTT	TTTACACACCA	GGAGGCTCAT	TTGCGCAGAC
	801	TGGCTACCCA	ATGTTTATCT	ATGATGCCCA	AAGGCACAA	TGGTTRATTA
	851	ATGGGGTATG	CGAACACGCG	ACCCTCATTA	TAGGAAAGAT	CAATGGCTTC
	901	CAGCTAGTTC	GTAAGAAGTT	GTCTTATCAT	GAATCTTTTG	TGCGAGATAC
	951	ACCATTCAGTA	TTCTACGAAC	CACATCAAAA	TGGGAATTAAT	TTTTTAAAGC
	1001	ACAAATPATA	TGGCGCAGGA	TAAGCTCGAT	CGACACATAC	ACACATTTCT
	1051	CTACCTTTATA	TTATPAAACG	ACGACCGATT	CAATTGGTAT	ATGTTTCTTA
	1101	CTACGAGACA	CGAAGAGAGC	ATGAGAGGAT	TGCTGTGAGT	GGGGTCAATC
	1151	GTATTGCGCC	CMAGCTGAT	ATGAGGAAAT	ATATTTCTTT	TATTTGACAA
	1201	GGAAATATCT	CGGAGGATCT	TGCGGCGGAT	CGCGGCGGAT	TGCGGCGGAT
	1251	TTTGTGTTAT	GGAGGTAAAT	TGCTTAAACT	CTCTTAAAC	ACGAAACGCT
	1301	GGGAGGCGCT	GGCGCTTCTA	ATCATGCTAT	CGCATACCTT	TACCTTGGAA
	1351	GTAAJGCGCG	TGGCAACGA	CGCGCTTCC	AAJATCGCGA	AGGGCAGGCT
	1401	GTGCTGTCAC	GCCAAAGGCG	ACCAACCAAGG	CTCGGTCAGC	TGCGGCGAGC
	1451	TGAATGTCAT	TTATAGATCAG	CAGGGCGATC	TACGAGGCAC	AAJACCAAGC
	1501	TTTATGTCAA	TGCGCTTGCT	CAGCAGGAGC	GGGACGGTGC	AACTGAATGC
	1551	CGATATPACG	TTCTAACCCC	ACCAACTCTA	TTTCCGCTTT	CGCGCGGGAC
	1601	TTTGTGGATT	GAACGGGCGT	TGCGTTTCTT	TCCACCGAT	TCAAAATACG
	1651	GATGAGGGGG	CGATGATTGT	CAACCAACAT	CAGACACAAAT	ACATCAACCT
	1701	TACCCATGCA	GGCAATAAAG	ATTAATCTAC	ACGCGGCAAT	AAACACACAT
	1751	TGGATAGCAA	AAAAGAAATT	GGCTACACAG	GTTCGTTTGG	CGAGAAAGAT
	1801	CGACACAAAA	CGACGGGCGC	GCTCAATCTG	AAITPACACG	CGGAGAGAGC
	1851	GGATCGGCGT	TTATGCTGTT	CGTCCGCGAG	AAATTTAAAC	GGCGAATATCA
	1901	CGCAACATCT	CGCAACATCT	TGCGGAGGCG	CGAGAGGAGT	ACGCAACATCA
	1951	TACATATATT	TACAGAGCGG	GTGTCGATCA	CGGCGAGAG	TCCCACAGAT
	2001	AGAAATCGTG	TGGGACGAGC	ATGTTGTCGA	CGCCACNTT	AGAGCGAAAT
	2051	ATCTCCATAT	TCCAGCGCGA	CAAGCGCTGC	TTTCCGCGAA	TGTTGCCAAA
	2101	GTGGAGAGGCT	ATTGGCATTT	AAGCATPAC	GCGGACAGAT	TGCTTGSGT
	2151	CGACCGGCTT	CAAGCCACCA	CACTGTCTAC	AGCTTCGAGT	TGAGCGGTTCA
	2201	TGACAGAGAT	TACGAAAGTA	ACCATTAACG	ACGNTAAAGT	GATTCGGTTC
	2251	TGCGACGAGA	CGCAGCTACG	AGGCATCTCA	CCGCTTGGCC	ATCAGCTGCT
	2301	TTTAAATCTC	ACAGGACTTG	CACATCTCAA	AGCAATCTTT	AGTGCAGGGC
	2351	GAGACGACGA	CTATAGCGTT	ACGCGCAAGC	CAACCACAAA	GGCGCAACCT
	2401	AGCGTCTGCG	CGAATGCCCA	CGACATATT	ATTCAAGCA	CATTAAACCT
	2451	CACACATCGC	CTGTCGACGA	ATGCTCTTAT	TAATCTAAGC	AAACACGCGC
	2501	TACAAATATCG	CGGTTCTGAG	TTTCTCCGAA	AGCTAATAGC	AAACGTAAAG
	2551	CATTGCGGAC	TCAACGACAG	TGTTCTCCTA	CGGCTAAGG	CGATATTTCA
	2601	CTTTTACATC	CTTTTACATC	ATTTAAAGG	ATTTAAAGG	CTTTTACATC
	2651	ATTATACAT	AAAGACAG	GTGAGAGCG	TGCGGCTGCT	CGGAGATTA
	2701	CGCAATTTAA	ACCTTGACAA	CGCCACNTT	ACATCTAAT	CGCGTCACTT
	2751	ACACATGTAG	ACGCGGCGCG	AAACCGCAGC	TGCGGCAAGT	CGCGCGCGCG
	2801	CGCGTGTGCG	CGGTTCCGCT	TATTCGSTTA	CGCGCGCAAT	TGCGCGAGAA
	2851	TCCGCTTTGCA	ACCGCTGTAC	GGTAAAGGCG	AAATTAAGC	GTACGGGAGC
70						

2901	ATTCGCGTTT	ATGTCGGAAC	TCTTCGSCTA	CGCGACGGC	AAATTGAAGC
2951	TGGCGGAAAG	TTCCGGAAGC	ACTTACACCT	TGGCTGTCAA	CAATACCGCG
3001	AACGAAACCG	TAACTCTCGA	GCAATTGAGC	GTAGTGGAA	GAAAGACAA
3051	CACACGCGTG	TCGGAATAATC	TTAATTTCAC	CCTGCAaaAc	gaacacgtcg
3101	atgcggcgccg	atgcgcttat	cagcttatcc	gaaagacgsg	cagcttcgcg
3151	CTKCAATATC	CGTCAAGA	ACAGAGCTT	TCGCAAAAC	TCGGCAAGG
3201	gggagaaACA	GagggcgccT	TGACGGCAAA	ACAGGCAcaA	CTTGCGCCCA
3251	AAcaaacggc	ggaaaaAGAC	AAcgcgcaaa	gccttgAcgc	gctgtatgcg
3301	gCggggcgca	atgccaccga	AAAGCGagaa	agtgtttcg	aaccgGCCCG
3351	GCAGGCGAGC	GGGGAATAtg	cgcgATTAT	GCAGGCGGAG	GAAAGAAAA
3401	AACGGGTGCA	GGCGGATAAA	GACACGCGCT	TGGCGAAACA	CGCGGAAGCG
3451	GAAACCGGCG	CGGCTACCAC	CGCCTTCCCG	CGCGCCCGCG	CGCGCCGCGC
3501	GGATTTCGCG	CAACCGCAGC	CCCAACCGCA	ACCCCAACCG	CAGCGCGACC
3551	TGATCAGCGG	TTATGCCAAT	AGCGGTTTGA	GTGAATTTTC	CGCCACGCTC
3601	AACAGCGTTT	TGCGCGTACA	GGAACGAATTG	GACCGCGTGT	TGCGCGAAGA
3651	CGCGCGCAAC	GCGGTTTGGG	CAGCGGCGAT	CGGGACACCC	AAACACTACC
3701	GTTCGCAAGA	TTTCCGCGCC	TACCGCAAC	AAACCGACCT	CGCGCAAACT
3751	GGTATGCAGA	AAAACCTCGG	CAGCGGCGCG	GTCCGCACTC	TGTTTTGSCA
3801	CAACCGGACCT	GGAACACCT	TCGACAGCG	CATCGGCAAC	TGCGGACGGC
3851	TTGCCACAGG	TGCGTTTTC	GGGCAATACG	GCATCGGACG	TTTCGACATC
3901	GGCATCAGCG	CGGGCGGGG	TTTATGTAAG	GGCAGCCTTT	CAGACGGCAT
3951	CAGAGGCCAA	ATCCGCGGCC	GGGTGCTGCA	TTACGGCATC	CAGGCAAGAT
4001	ACCGCGCAGG	TTTCGGCGGA	TTTCGGCATG	CAACGACAT	CGCGCAACG
4051	CGCATATTTC	TCGAAAAGC	GGATTACCGA	TACGAAAACG	TCAATATCGC
4101	CACCCGGGCG	CTTGCAATCA	ACCGCTACCG	CGCGGGCATT	AAGCGAGATT
4151	ATTCATTCAA	ACCGGCGCAA	CACATTTCGA	TCAGCGCTTA	TTTGAGCCGTG
4201	TCCTATACCG	ATGCGCGTTC	CGGCAAAAGTC	CGAAGCGCGC	TCATACCGCG
4251	CGTATTGGCG	CAGGATTTCG	GCAAAAACCG	CAGTGGGAAA	TGGGGCGTAA
4301	ACGCGCAAA	CAAGAGTTTC	ACGCTGTCCC	TCCACGCTGC	CGCGGCCAAG
4351	GGGCGCGCAAT	TGGAAGCGCA	GCACAGCGCG	GGCATCAAA	TAGGCTACCG
4401	CTGGTAA				

This is predicted to encode a protein having amino acid sequence <SEQ ID 654>:

1	MKTTDKRTTE	THRKAPKTGR	IRFSPAYLAI	CLSFGLPQA	RAGHTYFGIN
51	YQYIRDFAEN	KGKFAVGAAD	IEVYNKGEEL	VGKSMTKAFM	IDFSVVSRRG
101	VAALAGDOYI	VSVAHNGGYN	NVDFGAEGSN	PQDHFRSYQI	KVRNNYKAGT
151	NGHPYGGDYH	MPRLHKFVTD	AEPVEMTSYM	DGWKYADLNL	YDPRVRIGAG
201	ROYWRSDEDE	PNNRESSYHI	ASAYSWLVGG	NTFAQNSSGG	GTVNLGSEKI
251	KHSPYGLFT	GGSGDGSF	MEFYDAKQK	WLINGVLQGT	NPYIGKSNFG
301	QLVRKDFWFD	ELFAGDTHSV	FYEPHQNGKY	FFNDNNNGAG	KIDAKHKHYS
351	LPYRLKTRTV	QLFNVSLSSET	AREPVPYHAAG	GVNSYRRLIN	NGENISFIDK
401	KGKGLILTSN	INQGAGGLYF	EGNFTVSPKN	NETWQAGVGH	ISDGTVTWVK
451	VNGVANDRLS	KIGKGTLLVQ	AKGENGGSVS	VGDGKVLIDQ	QADDQGRKQA
501	FSEIGLVSGR	GTVQLNADNQ	FNFDKLYFGF	RGGRLLDNLGH	SLSFHRIGNT
551	DEGAMTVNHN	QKEATVPTIT	GNKDITPTGN	NNLDSKELI	AINSGWFGKED
601	ATTKINGELNL	NYFPEADRT	LLSGGTMLN	GNITCTZNGEL	FFSRRPTFHA
651	VNHLGSGWSK	MEGIPQGEIV	WDNDWIDRTE	KAENHFIQGG	QAVVSRNVAK
701	VEGDWHLSNH	AOAFVGVAPH	QSHITCTRSD	WTGLTSCTEK	TTTDDKVIAS
751	LSKTDVGRNV	SLADHAHLNL	TGLATFNGNL	VQAEFTRTIRL	RANATQNGNL
801	SLVNGAQAFT	NQATLNGTNS	ASDNASFNLS	NNAVQNSGLT	LSDNKANVNS
851	HSALINGVSL	ADKAVHFHEN	SRFTGKISGG	KDITALLHKDS	EWTLPSGTTEL
901	GNLNLNDNATI	TLNSAYRHDA	AGAQTGSAAAD	APRRRRSLRL	LSVTPTPTSAE
951	SRFNTLTVNG	KLNGQGTFRF	MSFLFGYRSG	KLKLAESESG	TYTLAVNNVTG
1001	NEPVLSLEQLT	VVEGKDNTPL	SENINLFTQN	EHVDAGAWRY	QLIRKDGFEFR
1051	LHNPKVEQEL	SDKLGKAGET	EALTAQQAQ	LAQAQQAQED	NAQSLDALIA
1101	AGRNATEKAE	SVAEPARQAG	GENAGIMQAE	EEKRRVQADK	DTALAKQREA
1151	ETRPATTATP	RARRARRDLF	QPCFQFQFQF	QRDLISRYAN	SGLSSESATL
1201	NSVFVARRN	AVWTSGRUT	KFRSSQRTFA	YRQGTDLQRI	
1251	GMQNLSSGR	VGTLFSHNR	GNTFDDGNG	CARLAHAAVF	GCYSIGRTHA
1301	GISAGAGFSR	GSLSDGIRGK	IRRRVLHYGI	QARYRAGFGG	FGIEPHIGAT
1351	RYFVQKADYR	YENVNIATPG	IAFNRYFRAGT	KADYSFKPAQ	HISITPYLSL
1401	SYTDAASGVK	RTRNVNTAIV	QDFGKTRSAE	WGYNAEKGF	TLISLHAAAAK
1451	GPQLEAGHSA	GIKLYRW*			

Underlined and double-underlined sequences represent the active site of a serine protease (trypsin family) and an ATP/GTP-binding site motif A (P-loop).

65 ORF1-1 and ORF1ng show 93.7% identity in 1471 aa overlap:

-372-

		10	20	30	40	50	60
	orf1-1.pep	MKTTDKRTTETHRKAPKTGRIRFSPAYLAICLSFGILPQAWAGHTYFGINVOYYRDFAEN					
5	orf1ng-1						
		10	20	30	40	50	60
	orf1-1.pep	70	80	90	100	110	120
10	orf1ng-1	KGKFAVGAKDIEVYNKKGELVGKSMTKAPMIDFSVSVSRNGVAALGQOYIVSVAHNGGYN					
		70	80	90	100	110	120
	orf1-1.pep	130	140	150	160	170	180
15	orf1ng-1	NVDFGAEGRNPDQHRFTYKIVKRNNYKAGTKGHPYGGDYHMPRLHKFVTDAPVEMTSMY					
		130	140	150	160	170	180
	orf1-1.pep	190	200	210	220	230	240
20	orf1ng-1	DGRKYIDQNNYPDRVRIGAGRQYWRSEDEDEPNNRESSYHIASAYSWLVGNTFAQNGSGG					
		190	200	210	220	230	240
	orf1-1.pep	250	260	270	280	290	300
25	orf1ng-1	GTVNLGSEKIKHSPYGLPTGGSGDGSMPFIYDAQOKWLINGVLQGNPYIGKSNGF					
		250	260	270	280	290	300
	orf1-1.pep	310	320	330	340	350	360
30	orf1ng-1	QLVRKDWFYDEIFAGDTHSVFYEPQNGKYSFNDNNGTGKINAKHEHNSLPNRLKTRTV					
		310	320	330	340	350	360
	orf1-1.pep	370	380	390	400	410	420
35	orf1ng-1	QLFNVSLSSETAREPVYHAAGVNSYRPLNNGENISFIDEGKGELIILTSNIQAGAGLYF					
		370	380	390	400	410	420
	orf1-1.pep	430	440	450	460	470	480
40	orf1ng-1	QGDFTVSPENNETWQAGGVHISEDSTVTWKVNGVANDRLSKIGKTLHVQAKGENQGSIS					
		430	440	450	460	470	480
	orf1-1.pep	490	500	510	520	530	540
45	orf1ng-1	VGDGTVILDQADQKQAFSEITGLVSGRGTVQLNADNQFNPDKLYFGFRGGRDLNIGH					
		490	500	510	520	530	540
	orf1-1.pep	550	560	570	580	590	600
50	orf1ng-1	SLSFHRIGNTDEGAMIVNHNQDKESTVTTGNKDIATTGNNNLSDKKELIAYNGWFGKD					
		550	560	570	580	590	600
	orf1-1.pep	610	620	630	640	650	660
55	orf1ng-1	TTKTNGRLNLVYQPAADRITLLSGGTNLNGNITQTNGKLFFSGRPTPHAYNHLNDHWSQ					
		610	620	630	640	650	660
	orf1-1.pep	670	680	690	700	710	720
60	orf1ng-1	MEGIPQGEIVWDNDWIRTFKAENFHTQGGQAVVSRNVAKVGDWHLNSHAQAVFGVAPH					
		670	680	690	700	710	720

730 740 750 760 770 780
orf1-1.pep QSHTICTRSDWTGLTNCVEKTTIDDDKVASLTKTDISGNVDLADHAHLNLTGLATLNGNL
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
orf1ng-1 QSHTICTRSDWTGLTSCFEKTTIDDKVASLSKTDIRGNSVLADHAHLNLTGLATLNGNL
730 740 750 760 770 780

790 800 810 820 830 840
orf1-1.pep SANGDTRTYTVSHNATQNGNSLVGNAQAFTNQATLNGMTSASGNASFNLSDHAVONGSLT
| | | : | : | : | : | : | : | : | : | : | : | : | : | : |
orfing-1 SAGGDTHYTVTRNATQNGLSLVGNAQAFTNQATLNGMTSADSNASFNLSSNAVONSGLT
790 800 810 820 830 840

15

	850	860	870	880	890	900
orf1-1.pap	L S G N A K A N V H S A L N G V S I A D K A V F H F E S R R T G Q I S G G K D T A L H L K D S E W T L S G T L					
orf1ng-1	L S D N A K A N V H S A L N G V S I A D K A V F H F E S R R T G K I S G G K D T A L H L K D S E W T L S G T L					
	850	860	870	880	890	900

```

20      910      920      930      940      950      960
orf1-1.pap  GNLNLNATITILNSAYRHDAAAGATQGSADAPRRRRSRRLSVPTTSPVESFNTLT
|||||
orfing-1    GNLNLNATITILNSAYRHDAAAGATQGSADAPRRRR---RLSLSVPTTSPAESFNTLT
      910      920      930      940      950

```

25

```

          970      980      990      1000      1010      1020
orf1-1.pep  VNGKINGGQTFRFMSELFGRVSDKIKLIAESSEGTYYTLAVNNTGNEPVSLEQLTVVEGKDN
|||||
orf1ing-1   VNGKINGGQTFRFMSELFGRVSDKIKLIAESSEGTYYTLAVNNTGNEPVSLEQLTVVEGKDN
          960      970      980      990      1000      1010
```

[illegible]

```

                                1080   1090   1100   1110   1120
orf1-1.pep  ----EAKRQAEKDNAQSLDLIAAGRDVAEKTESVAEPAQAAGGENVNGIMQAEEEKRQV
              ||:|||||:|||||:|:|:|||||:|:|:|||||:|:|:|||||:|:|:|||||:|:|:
40 orf1ng-1  QAQAARQAQAEKDNAQSLDLIAAGRNATEKASVAEPAQAAGGENAGIMQAEEEKRQV
              1080   1090   1100   1110   1120   1130

```

45

orf1-1.pep	1130	1140	1150	1160	1170	1180
	ADKDTALAKQREATRPATTAFPRARRARRDLPLQPPQPPQPPQDRLISKYANGSLSEFS					
orf1ng-1	ADKDTALAKQREATRPATTAFPRARRARRDLPLQPPQPPQPPQDRLISKYANGSLSEFS					
	1140	1150	1160	1170	1180	1190

50 orf1-1.pep ATLMNVFAVQDELDRVFAEDRRNAVWTSIGRIDTKHYRSQDFRAYRQQTDLRQIGMQKNLG
orf1ng-1 ATLMNVFAVQDELDRVFAEDRRNAVWTSIGRIDTKHYRSQDFRAYRQQTDLRQIGMQKNLG

55

	1250	1260	1270	1280	1290	1300
orf1-1.pep	SGRVGILFSHNRTENTPDDGIGNSARLAHGAVFGQYGIDRFYIGISAGAGFSSGSLSDGI					
orf1ng-1	SGRVGILFSHNRTGNTPDDGIGNSARLAHGAVFGQYIGRFDIGISAGAGFSSGSLSDGI					
	1260	1270	1280	1290	1300	1310

65

1310 1320 1330 1340 1350 1360
orf1-1.pep GGGKRRRVLYHGIQARYRAGFGGFGIEPHIGATRYFVQKADYRYENNVIATPGLAFNRYR
orf1ng-1 RGGKRRRVLYHGIQARYRAGFGGFGIEPHIGATRYFVQKADYRYENNVIATPGLAFNRYR
1320 1330 1340 1350 1360 1370

1370 1380 1390 1400 1410 1420
orf1-1.pep AGIKADYSFKPAQHISITPYILSLSYTDAASGKVRTRVNTAVLAQDFGKTRSAEWGVNAEII
70 orf1ng-1 AGIKADYSFKPAQHISITPYILSLSYTDAASGKVRTRVNTAVLAQDFGKTRSAEWGVNAEII
1380 1390 1400 1410 1420 1430

		600	610	620	630	640	650
	orf1ng-1.pep	EKDATKTINGRLNINYPQEEADRTLLSGGTNLNGNITQTNGKLFFSGRTPHAYNHIGSG					
5	p45387	ETDRKNHNGRLNLIYKPTTEDRTLLSGGTNLKGDITQKGLFFSGRTPHAYNHINLR	570	580	590	600	610
						620	
		660	670	680	690	700	710
	orf1ng-1.pep	WSKMEGIPQGEIVVDNDWIDRTTKAENFHIGGSQAVVSRNVAKVEGDWHLNSNHAQAVFGV					
10	p45387	WSKMEGIPQGEIVVDHDWINKTFKAENFIKGGSAVVSNNVSSIEGNWTVSSNNATATFGV	630	640	650	660	670
						680	
		720	730	740	750	760	770
	orf1ng-1.pep	APHQSHTICTRSDWTGLTSCTEKTIIDDKVIASLSKTDIRGNVSLADHAHLNLTGLATLN					
15	p45387	VFNQNTICTRSDWTGLTTCQKVDLTDTKVINSPKTKINGSINLTDNATNAVVKLAKLN	690	700	710	720	730
						740	
		780	790	800	810	820	830
	orf1ng-1.pep	GHLGAGDTHYTVTRNATQNGNLVLGNAQATTNQAATLNGNTSASDNASFNLSNNAVQNG					
20	p45387	GNVTL-----TNHSQFTLSNNATQIG	750			760	770
		840	850	860	870	880	890
	orf1ng-1.pep	SLTSLDNAKANVSHSALGNVSLADKAVFHFENSRFTGKISGGKDTALHLKDSWTLPSG					
25	p45387	NRLSDNSTATVDNANLNGMVLTDQAQSLKNSHFSGIQGDKGTVTTLNENATWTFPSD	840	850	860	870	880
		900	910	920	930	940	950
	orf1ng-1.pep	TELGNLNLDNATITLNSAYRHDAAGAQTGSAADAPRRRSRRLSVTPPTSARESFNTLT					
30	p45387	TTLQNLTLNNSTITLSAY-----SASSNNTPRRS---LETFETTPTSARESFNTLT	900	910	920	930	940
		960	970	980	990	1000	1010
	orf1ng-1.pep	VNGKLGQGTFRFMSELFGYRSGKLKLAESSEGTYYTLAVNNTGNEPVFSLEQLTVVEGKDN					
35	p45387	VNGKLGQGTFTQFTSSLFGYKSDKLKLSNDAEGDYILSVRNTGKEPETLEQLTLVESKDN	960	970	980	990	1000
		1020	1030	1040	1050	1060	1070
	orf1ng-1.pep	TFLSENLFNLQNEHVDAGAWRYQLIRKDCGFRLNHNVEQELSDKLGKAGETEALATK					
40	p45387	QPLSDKLKFTLENDHVDAGALRYKLVKNDGEFRLNHNPIKEQLHNDLVRAEQARTLEAK	1020	1030	1040	1050	1060
		1080	1090	1100	1110	1120	1130
	orf1ng-1.pep	QAQLAAKQQAEDNAQSLDALIAAGRNAT-EKASVAEPAQAGGENAGIMQAEEEKKRV					
45	p45387	QVEPTAKTQTGEKPVSRRAARAAPDPTLPDQSLNLALEAKQAE-LTAETQKSKAKTKV	1080	1090	1100	1110	1120
		1140	1150	1160	1170	1180	1190
	orf1ng-1.pep	QADK---DTALAKQREARETREATTAFFARRARRRD-LFQFPQFPQFPQFQDRLISRYANGS					
50	p45387	RSKRAVFSDDLPLDKSLFALEAALEVIDAPQOSEKDRLAQEAEEKQ-KKQDLISRYNSNA	1140	1150	1160	1170	1180
		1200	1210	1220	1230	1240	1250
	orf1ng-1.pep	LSEFSATLNSVFVQDELDRVFAEDRRNAVWTSGIROTKHYRSQDFRAYRQQ-TDLRQIG					
55	p45387	LSELSATVNSMLSVQDELDRFLVDQAQSAVWNIQAQKRYVDSDFRAYQQQKTNLRQIG	1200	1210	1220	1230	1240
		1260	1270	1280	1290	1300	1310
	orf1ng-1.pep	MQKNLGSGRVGLFSHNRTGMTFTDDGIGNSARLAHGAVGQYIGRFDIGISAGAGFSSG					
60	p45387	VQKALANGRIGAVFSHRSRSDNTFDEQVKMHAETLTMMSGAGYQWGDLPQGVNVTGISAS	1260	1270	1280	1290	1300
		1180	1190	1200	1210	1220	1230

-376-

		1320	1330	1340	1350	1360	1370
orf1ng-1.pep		SLSDGIRKIRRRVLHYGIQARYRAGFGGFGIEPHIGATRYFVQKADRYENVNATPGL					
		1240	1250	1260	1270	1280	1290
5	p45387	1240	1250	1260	1270	1280	1290
		1380	1390	1400	1410	1420	1430
orf1ng-1.pep		AFNRYRAGIKADYSFKPAQHISTTPTSLSYTDAASGKVRTRVNTAVLAQDFGKTRSAEN					
		1300	1310	1320	1330	1340	1350
10	p45387	1300	1310	1320	1330	1340	1350
		1440	1450	1460	1470		
orf1ng-1.pep		GVNAEIKGFTLSLHAAAAGFQLEAQSAGIKLGYRWX					
		1360	1370	1380	1390		
15	p45387	1360	1370	1380	1390		

Based on this analysis, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

20 Example 78

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 655>:

	1	..AAGGTGTGGC	AATTTGTGCA	AGA.CCGCTG	CGTGCCGCTG	TGCCTGCCGA
	51	CAGTTTGTAA	CGACCCGCGC	AAAAATTGAA	CCTGTTTAAG	CGCGGTGCGG
	101	CAACCATTTT	GTTTATGAA	GATCAAAATG	TGCTCAAAAG	TTTGAGGAG
25	151	CAGTTCCTTG	CTTATGCCGC	TAACTTCCCC	GTTTGGGCGG	ATCAGGCAAA
	201	CGCGATGGTG	CAGTATGCCG	TTTGACGAC	ACTTGGCGCG	GTCCGCGTAG
	251	GTGCAAACT	GCAACATTAC	AATCCCTTGC	CCGATCGGGC	GATTGCCAAA
	301	CGGTGGAATA	TCCCGAATAA	CTGTTTGTTC	CGCGCACAAA	TGTTTATCGG
	351	CGGTATTGAA	GGGGCGGCGC	GTGAAAGAC	CTTTGAACCC	GTTCAGAAC
30	401	GTTCGAAAGT	GTTCGGCGCA	TAA		

This corresponds to the amino acid sequence <SEQ ID 656; ORF6>:

	1	..KVNQFVEXPL	RAVVPADSTF	PTAQKLNLFK	AGAATILFYE	DQNVVKGLE
	51	QFPAYAAANF	VWADQANAMV	QYAVWTTLLA	VGVGANLQHY	NPLPDAAIAK
	101	AWNIPENWLL	RAQMVGIGIE	GAAGKTFEP	VAERIKVFGA	*

35 Further sequence analysis revealed a further partial DNA sequence <SEQ ID 657>:

	1	..CTGCGTGCCG	TCGTGCTCTG	CGACAGTTT	GAACGACCGC	CGCAAAAATT
	51	GAACCTGTTT	AAGGCGGGTG	CGGCAACCAT	TTTGTTTTAT	GAAGATCAAA
	101	ATGTCGTCAA	AGGTTTGCAG	GAGCAGTTCC	CTGCTTATGC	CGCTAACTTC
	151	CCGCTTTGGG	CGGATCAGGC	AAACGCGATG	GTGCAATATG	CCGTTTGGAC
40	201	GACACTTGCC	CGGCTGGGCG	TAGGTGCAAA	CCTGCAACAT	TACAATCCCT
	251	TGCGCGATGC	GGCGATTGCG	AAAGCGTGGG	ATATCCCCGA	AAACTGGTTG
	301	TTGGCGGCAC	AAATGOTTAT	CGGCGGTATT	GAGGGGGCGG	CAGGTGAAAA
	351	GACCTTTGAA	CCGCTTGCGA	AACTTTGAA	AGTGTTCGCG	GCATAA

This corresponds to the amino acid sequence <SEQ ID 658; ORF6-1>:

	1	..LRVAVPADSF	EPTAQKLNLF	KAGARTILFY	EDQNVVKGLE	EQFPAYAAANF
	51	PVWADQANAM	VQYAVWTTLLA	AVGVGANLQHY	YNPLPDAAIA	KAWNIPENWL
	101	LRQMVGIGIE	EGAAGKTFE	PVAERIKVFG	A*	

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

50 ORF6 shows 98.6% identity over a 140aa overlap with an ORF (ORF6a) from strain A of *N.meningitidis*:

5
10
15

orf6.pep
orf6a
orf6.pep
orf6a
orf6.pep
orf6a

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
83

20	1	ATGACCGCGTC	AATCTTCGCA	ACAGGTCGGC	GAAGGCGGAC	GTTCACATTA
	51	TCGTGTAATTC	AAAATCTCGC	CGCTCGGCAAA	GTCCAAATAG	
	101	TGCAAGCACGC	CGCTTTTGAC	ACACCTCTCT	CGCTCAATTC	CCAACTCGGAT
	151	CGTGTGTGGCG	TGCTGTTTGG	CGACATGATG	GATAGAGGTG	GGCAATTTGGT
	201	CGAAGACGCG	TCGCTGTGGC	TTCGTCGCGC	CGACAGCTTTT	GACCCGACCG
25	251	CGCAAAATAT	GAACCTGTGT	AGGCGGGTGG	CGGCACATAT	TTTGTTTATAT
	301	GAGATCAAA	ATGTCGTGCA	AGGTTTCGAG	GAGCACTTCC	TGCTTTATGCG
	351	CGGCAATCTT	CCGCTTTGGG	CGGACGACGC	GACGCGGATG	TGCGATGATG
	401	CGGTTTGGAC	GATCATCTGC	GGGCTGTGGC	TAGGTCGAAA	CCTGCACATAT
	451	TGCAATCTGAT	GGATCTGATG	GGGCTGTGGC	ATATCGGATG	GAGGCGGACG
30	501	CAAGTGTGAA	TGTCGCGACG	CAATGTTTAT	ACGCGTGTAT	GAGGCGGACG
	551	CGGTGTGAAA	GACCTTTGAA	CCAGTTTCAG	CAGCTTTGAA	AGTGTTCGGC
	601	GCATAA				

35 1 MTRQSLQQA ESRRSIYSIN KNLVPGKDEI VQIVEHAVLH TPSSFNSQSA
51 EDVNVFGEEH DKWVQVEDA IRVAVPQAD EPTAQKLNLF KAGAAATILFY
101 EQQNVVVKGLQ EQPFAANF PWAQDAQNM VQYAVWTLTA AVGGANGLKFG
151 YNPLPDAATA KAWNIPENWL LRAQMVGIGI EGAAGEKTFE PVAERLKVFG
201 A*

40 orf6a.pep TFSSFNSQSRVVVLFGEEHDKVWQFVEDLARVVPADSFEP⁵⁰TQA⁶⁰KL⁷⁰IL⁸⁰FKA⁹⁰GAAT¹⁰⁰IL¹¹⁰IL¹²⁰Y
orf6-1 LRAVVPADSFEP¹⁰TQA²⁰KL³⁰IL⁴⁰FKA⁵⁰GAAT⁶⁰IL⁷⁰IL⁸⁰Y

45 orf6a.pep EDQNVVKG¹¹⁰LQEQ¹²⁰F¹³⁰YAA¹⁴⁰NFP¹⁵⁰WADQ¹⁶⁰ANAM¹⁷⁰VQY¹⁸⁰AV¹⁹⁰W²⁰⁰T²¹⁰LAA²²⁰VG²³⁰GAN²⁴⁰LQ²⁵⁰HN²⁶⁰PL²⁷⁰DA²⁸⁰AA²⁹⁰IA³⁰⁰
orf6-1 EDQNVVKG¹¹⁰LQEQ¹²⁰F¹³⁰YAA¹⁴⁰NFP¹⁵⁰WADQ¹⁶⁰ANAM¹⁷⁰VQY¹⁸⁰AV¹⁹⁰W²⁰⁰T²¹⁰LAA²²⁰VG²³⁰GAN²⁴⁰LQ²⁵⁰HN²⁶⁰PL²⁷⁰DA²⁸⁰AA²⁹⁰IA³⁰⁰

50 orf6a.pep KANNIPEN¹⁷⁰NWL¹⁸⁰LRAQ¹⁹⁰W²⁰⁰IIGG²¹⁰E²²⁰GA²³⁰AGE²⁴⁰K²⁵⁰TF²⁶⁰PAER²⁷⁰LK²⁸⁰VF²⁹⁰GAX³⁰⁰
orf6-1 KANNIPEN¹⁷⁰NWL¹⁸⁰LRAQ¹⁹⁰W²⁰⁰IIGG²¹⁰E²²⁰GA²³⁰AGE²⁴⁰K²⁵⁰TF²⁶⁰PAER²⁷⁰LK²⁸⁰VF²⁹⁰GAX³⁰⁰

55 orf6a.pep
orf6-1

ORF6 shows 95.7% identity over a 140aa overlap with a predicted ORF (ORF6ng) from *N. gonorrhoeae*;

-378-

orf6.pep		KVWQFVEXPLRAVVPADSFEPTAQLNLFK	30
orf6ng	SNVLSMSNPTVLRMGLPLYIASLRGAIYK	WQFVEDALRAVVPADSFEPTAQLKLFK	64
5 orf6.pep	AGAATILFYEDQNVVKGLEQFPAYANFPVWADQANAMVQYAVWTTLAAVGVCANLQHY		90
orf6ng	AGAATILFYEDQNVVKGLEQFPAYANFPVWADQANAMVQYAVWTTLAAVGAGANLQHY		124
10 orf6.pep	NPLPDAALAKAWNIPENWLLRAQMVGIGIEGAAGEKTFEPVAERLKVFGA		140
orf6ng	NPLPDVAALAKAWNIPENWLLRAQMVGIGIEGAAGEKVFEPVAERLKVFGA		174

The complete length ORF6ng nucleotide sequence <SEQ ID 661> was identified as:

15	1	ATGGCCGTTG	CGTCAAATGT	CAGCTTGGAT	ATGTCCAATC	CTACGGTGT	1
	51	ACGCATGGGA	TTACCCCTTAT	ATATTGCGTC	CCTAAGAAGG	GGCGCAATAT	
	101	ATAAGGTGTG	GCAATTTTGT	GAAGACGCGC	TGCGTGCGGT	CGTGCCTGCC	
	151	GACAGTTTGT	AACCGACCGC	GCAAAATTTG	AAGCTGTGTTA	AGGCGGGCGC	
	201	GGCAACCAT	TGTTTTATG	AAGATCAAAA	TGTGCTCAAA	GGTTTCAGG	
	251	AGCAGTTCCC	TGCTTATGCC	GCCACCTTTC	CGTTTGGCG	GGACGAGCG	
20	301	AACGCTATGG	TACAGTATGC	CCTCTGGAGC	ACACTTGGCG	CGCTCGGTGC	
	351	AGGTGCAAT	CTGCAACATT	ACAACCCCTT	GGCGATGTG	CGGATTGCTA	
	401	AAGCGTGGAA	TATTCGCCGA	AACCTGGCTGT	TGCGCGCGCA	AATGTTATC	
	451	GGTGTATTG	AAGGGGcgcc	aggtgaaaaa	gtctttgaac	CCGTTGCGga	
	501	acg:tttgAAA	GTGTTGCGCG	CATAA			

25 This encodes a protein having amino acid sequence <SEQ ID 662>:

30	1	MAVASNSVLD	MSNPTVLRMG	LPLYIASLR	GAIFYKWQFV	EDALRAVVP	
	51	DSFEPTAQL	KLFKAGAA	ILFYEDQNVK	GLQEQFPAYA	ANFPVWADQA	
	101	NAMVQYAVT	TLAAGVAGAN	LQHYNPLDV	AIKAWNIP	NWLLRAQMVI	
	151	GGIEGAAGEK	VFEPVAERLK	VEGA*			

ORF6ng and ORF6-1 show 96.9% identity in 131 aa overlap:

	orf6-1.pep				10	20	30
					LRVVVPADSFEPTAQLNLFKAGAAITILFY		
35	orf6ng	PTVLRMGLPLYIASLRGAIYKWQFVEDALRAVVPADSFEPTAQLKLFKAGAAITILFY					
		20	30	40	50	60	70
	orf6-1.pep		40	50	60	70	80
40		EDQNIVVKGLEQFPAYANFPVWADQANAMVQYAVWTTLAAVGVCANLQHYNPLPDAIA					
	orf6ng	EDQNIVVKGLEQFPAYANFPVWADQANAMVQYAVWTTLAAVGAGANLQHYNPLPDVAIA					
		80	90	100	110	120	130
	orf6-1.pep		100	110	120	130	
45		KAWNIPENWLLRAQMVGIGIEGAAGEKTFEPVAERLKVFGAX					
	orf6ng	KAWNIPENWLLRAQMVGIGIEGAAGEKVFEPVAERLKVFGAX					
		140	150	160	170		

50 It is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 79

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 663>

55	1	..GGCTACAAC	ACCTGTTCGC	GGCGGCGAG	GGCATCGCA	ACTACCAAT	
	51	CAACGGCATC	CCGCTTGCG	ACGCGCTGG	CGATACGGG	CAATCGAAC	
	101	ACCGCCGCT	ATGACCGCT	AGAGCTGTG	CGCGGTGG	CGGCGCTGT	
	151	GGACGGCAG	GGGAGCCTT	CCGCCCGGT	CAATCTGGT	CGCAACGCC	
	201	TGACCCGCCA	CCCATTTT	GAAGTCCGC	CCGAGCGGG	CAACCGCAA	


```

251 CATTTCGGGC TGGACGCGGA CGTATCGGGC AGCGTGAACA CGGAAG.cyc
301 xCTGCGCGcG CGCCTGTGTT CCACCTTCGG ACGGCGCGAC CGTGTGGCGC
351 GCGCGGAAAGC GAGCCGskAT GCGGAACCTT ACGGCATTTT GGAATACGAC
401 ATCGCACCGC AAACCCGCGT CCACGCAcGc ATGGACTACC AGCAGCGGAA
451 AGAAACCGCC GACGCGCGCG TCAGcTACGc CGTGTACGAC AGCCAAGGTT
501 ATGCCACCGC CTTCGCGCGG AAAGCAACAC CGGCGACAAA TTGGCGCAAC
551 AGCCACGACC GTGCGCTCAA CCGTTCGCGC GGCATCGAAC ACCGTTTCAA
601 CCAAGACTGG AACTCAAAG CGGAATACGA CTC..

```

This corresponds to the amino acid sequence <SEQ ID 664; ORF23>:

```

10 1 ..GNYNLFARGs RIANYQINGI FVADALADTG NANTAAYERV EVVRGVAGLL
51 DGTGEPsATV NLVRKRLTRK PLFEVRAEAG NRKHFGLDAD VSGSLNTEXX
101 LRGLRVSTFG RGDsWRRRRR SRXAEYLGI EYDIAPOTRV HAXMDYQQAk
151 FTADAPLSYA VYDSQGYATA FGPKNPATN WANSHHRLN LFAGIEHREN
201 QWKLKAEYD Y..

```

15 Further work revealed the complete nucleotide sequence <SEQ ID 665>:

```

1 ATGACACGCT TCAAAATATTC CTTCTGTGTT GCGGCCCTGT TCGCGGTGTA
51 CGCGCAGGCT GATGTTTCTG TTTCAGAGCA CCCCACACCT GAGGAAGAAG
101 CTGAATTTGCC GACCATCACG GTTACCGCGC ACCGCAACCG GAGTTCACAC
151 GAGCGGTACA CTGTTTCGCG CAGCGACACG CCGGTGGGCG TGCCCATGAC
201 CTFGGCGGAA ATCCGCGAGA GCGTCACGCT CAGTCATATG CAAACAATGC
251 GCGACCAAAA CATCAAACG CTGCAACGCG CCCTGTGTGA GCGACCGGCG
301 ACCAGCGCGC AGATTTAACG CTCGACCGCG CGGCGGTACA ACTACTGTT
351 CGCGCGCGCG AGCGGCATCG CCAACTACCA AATCAACGCG ATCCCGTTG
401 CGGACGCGCT GCGCGATACG GGCAATGCCA ACACCGCGCG CTATGAGCGC
451 GTAGAAGTCT TGCGGGGCGT GGGGGGGGCT GTGGACGGCA CGGCGAGCGC
501 TTCCGCGCAC GTCAATCTGG TGCGCAAAAG CCGTACCGCG AAGCCATTGT
551 TTGAAGTCCG CGCGCAAGCG GGCAACCGCA AACATTTGCG GCTGGACGCG
601 GACGTATCGG CGACGCTGAA CACCGAAGCG ACGCTGGCGG CGCCGCTGCT
651 TTCCACCTTC GGACGCGCGC ACTCGTGGCG GCGGGGGGAA CGGAGCGCGC
701 ATGCGCGAAG CTACGGGACT TTGGAATACG AATGCGGACC GCACACCGCG
751 GTCCACGACT CATCGGACTA CACGAGCGCG AAGAAGACCG CGGACGCGCG
801 GCTCAGCTAC GCGCTGTACG ACAGCAACGCG TTATGCCACC GCTCTGGCG
851 CGAAGAGACA CCGCGCCACA AATTGGGGCA ACAGCGCCCA CCGTGGCGCT
901 AACCTGTTCG CCGCGCATGA ACACCGCTTC ACCAAGCACT GGAACTCAA
951 AGCGGAATAC GACTACACCC CGACCGCGCT CCGCGACGCC TACGGCGTAG
1001 CAGGCGTGCT TTGCATCGAC CACAACACCG CGCGCACCGA CCGTATTCCC
1051 GGTATTATGGC ACGCGCAGCC GCGCACCCAC AGCGCGAGCG TGCTATTGAT
1101 CGGCAAAATAC CGCTGTGTCG GCGCGGAACA CGATTTAATC GCGGCTATCA
1151 ACGGTTACAA ATACGCGCAG AACAATATCG GCGAAGCGAG CATCATCCCC
1201 AACGCCATTG CCAACGCGCTA CGAATTTTCC CGCACGGGTG CCTACCCGCA
1251 GCGTGCATCG TTTGCCCAAA CCAATCCCGA ATACGGGACC AGGCGCGAAA
1301 TCGGCGGCTA TCTGCGCACG CGTTTCGCGG CGCGCGACAA CCTTTCGCTG
1351 ATTTTGGGCG GACGATACAC CGGTATCCCG ACCGCGAGCT ACAGACGCGC
1401 CACACAAGCG ATGACCTATG TGTCGCGCAA CCGTTTCACC CCTACACAG
1451 GCATCGTGTG CGACCTGAGC GCAACCTGCT CTCTTACGCG CTGCTACAG
1501 AGCTGTCTCG TCCGCGCAAT CGAAAAGAGC GAACAGGCGA GCTACTGAAA
1551 ACCCGTAACC GGCACAATCT TGGAAAGCGG CATCAAGGCG GAATGCGCTG
1601 AAGCGCGTCT GAACGATCC CGCGCGCTGT ACCGCGCGCG TAAAAACAAC
1651 CTCGCCACCG CAGCAGGAGC GCACCGGAGC GGCACACCT ACTACGCGCG
1701 CGCCAAACCA GCCAAACACC ACGGCTGGGA AATCGAAGTC GCGCGCGCGA
1751 TCACGCCGGA ATGGCAGATA CAGGCGAGTT ACAGCCAAAG CAAAACCGCG
1801 GACCAAGAGC GACGCGGCTT GAACCCGAGC AGCGTACCGC AACCGAGCTT
1851 CAAACTCTTC ACTGCGCTACC ACTTTGCCCG CGAAGCGCCC AGCGGCTGGA
1901 CCAATCGCGC AGCGGTGCGC TGGCAGAGCG AAACCCACAC GCACCGTGGC
1951 ACGTCTCGCA TCCCAACACC CGCGGCCAAA GCGCGCGCGC CGCAACAGAG
2001 CGCGCAAAAG GCGTACGCGC TCGCGGACAT CATGGCGCGT TACCGCTTCA
2051 ATCCGCGCGC CGAATCTGCG CTGAAGCTGG ACAATCTGTT CACAAACAC
2101 TACCGCACCC AGCGCGAGCG CCAACGCTAC GGGCGACTCG GACAGTGAA
2151 CGCGCGGCTT ACCTATCGGT TTAAATTA

```

60 This corresponds to the amino acid sequence <SEQ ID 666; ORF23-1>:

```

1 MTRKYSLLF AALLEVYAGA DVSVSDDPKP QESTLEPTIT VTADRTASSN
51 DGIVTSGTHT PLGLEMTLRE IFQSVSVITS QMRDQNIKT LDRALLGATG
101 TSRIQYGS DR AGYNLFLARG SRIANYQING FVADALADTG GNANTAAYER
151 VEVRGVAGL LDGTGEPsAT VNLVRKRLTR PLFEVRAEAG NRKHFGLDAD
65 DVSGLNTEG TLRGLRVSTF RGDsWRRRRR RSRDAEYLGI LEYDIAFOTR

```

-380-

251 VHAGMDYQQA KETADAPLSY AVYDSQGYAT AFGPKDNFAT NWANSRHRAL
 301 NLFAGIEHRF NQDWKLKAEY DYTRSRFRQP YGVAGVLSID HNTAATDLTP
 351 GYWHADPRTH SASVSLIGKY RLFGRHDLI AGINGYKYAS NKYGERSIIP
 401 NAIPNAYEFS RTGAYPQPAS PAQTIPQYGT RRQIGGYLAT RFRADNLSL
 451 ILGGRYTRYR TGSYDSRTQG MITYVSANRFT PYTGIVFDLT GNLSLYGSYS
 501 SLFVFSQSKD EHGSLYLPVT GNNLEAGIKG EWLEGRNAS AAVYRARKMN
 551 LATAAGRDFS GNTYFRAANQ ARTHGWEIV GKRITPEWIO QAGYSQSKTR
 601 DQDGSRLNPD SVPEPSFKLF TAYHFAPEAZ SGWITGAGYR WQSETHDPDA
 651 TLRIENPAAK ARAADNSROK AYAVADIMAR YRNPRAELS LNVNLFNKH
 701 YRTQPDHRSY GALRTVNAAF TYRFK*

Computer analysis of this amino acid sequence gave the following results:

Homology with the ferric-pseudobactin receptor PupB of *Pseudomonas putida* (accession number P38047)

ORF23 and PupB protein show 32% aa identity in 205aa overlap:

Orf23 6 FARGSRINYQINGIPVADALADTGNANTAAAYERVEVVRGVLDDGTGEP SATVNLRK 65
 ++RG I NY+++G+P + L D + + A ++RVE+VRG GL+ G G PSAT+NL+RK
 PupB 215 WSRGFAIQNYEVDGVPTSTR-LDYSQSMAMFORVEIVRGATLSTGSGMGNPSATINLRK 273
 Orf23 66 RLTRKPLFEVRAEAGNRKHFGLDADVSGSLNTEKXLRGLVSTFKXXXXXXXXXXXXXAE 125
 R T + + EAGN +G DVS G L +RGR V+ +
 PupB 274 RPTAEQAASITGEAGNWDYRGTFGDFVSGELTETGIRGIRFVADYKTEKAWIDRYNQSQSL 333
 Orf23 126 LYGILEYDIAPQTRVHAXMDYQQA KETADAPLSYAVYD--SQGYATAFGPKDNFATNWN 183
 +YGI E+D+ + T + Y + D+EL + S G T N A +W+
 PupB 334 MYGITEFDLSEDTLLTVGFSY--LASDIDSLRSGLPTRFSTGERTNLKRSLNAEFDWSY 391
 Orf23 184 SHRALNLFAGIEHRFNQDWKLKAE 208
 + H + F IE + W KE
 PupB 392 NDHEQTSFFTSIEQQLGNGWSGKIE 416

30 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF23 shows 95.7% identity over a 211aa overlap with an ORF (ORF23a) from strain A of *N. meningitidis*:

orf23.pep 10 20 30
 GYNLYFARGSRINYQINGIPVADALADTG
 35 orf23a QMRDQNIKALDRALLQATGTSRQIYGS DRAGYNLYFARGSRINYQINGIPVADALADTG
 90 100 110 120 130 140
 orf23.pep 40 50 60 70 80 90
 NANTAAYERVEVVRGVLDDGTGEP SATVNLRKRLTRKPLFEVRAEAGNRKHFGLDAD
 orf23a NANTAAYERVEVVRGVLDDGTGEP SATVNLRKRLTRKPLFEVRAEAGNRKHFGLGAD
 150 160 170 180 190 200
 orf23.pep 100 110 120 130 140 150
 VSGSLNTEKXLRGLVSTFGRGDSWRARRSRKAELEYGILEYDIAPQTRVHAXMDYQQA
 orf23a VSGSLNTEKXLRGLVSTFGRGDSWRARRSRKAELEYGILEYDIAPQTRVHAXMDYQQA
 210 220 230 240 250 260
 orf23.pep 160 170 180 190 200 210
 ETADAPLSYAVYDSQGYATAFGPKDNFATNWNASHRALNLFAGIEHRFNQDWKLKAEYD
 orf23a ETADAPLSYAVYDSQGYATAFGPKDNFATNWNASHRALNLFAGIEHRFNQDWKLKAEYD
 270 280 290 300 310 320
 orf23.pep Y
 I
 60 orf23a YTRSRFRQPYGVAGVLSIDHNTAATDLIPGYWHADPRTHSASVSLIGKYRLFGREHDLIA
 330 340 350 360 370 380

The complete length ORF23a nucleotide sequence <SEQ ID 667> is:

1 ATGACACGCT TCAAAATATTC OCTGCTGTTT GCCGCCCTGT TGCCCGTGTA
 51 CGCGCAGGCC GATGTTTCTG TTTACAGACG CCCCAAACCG CAGGAAAGCA
 101 CTGAATTGCC GACCATTACC GTTACCGCGC ACCGCACGCG GAGTTCACAC
 151 GACGGCTACA CTGTTTTCGGC CAGCAGACAC CGCGCTGGGC TGCCCATGAC
 201 CCTSGCGGAA ATCCCGGAGA GCGTCAGCGT CATCAACATG CAACAAATGC
 251 CGACGCAAAA CATCAAGCGG CTGACGACGC CCTCTGTGCG GCGCAGCGGC
 301 ACCGCGCGCC AGATTATACGG CTCGCACCGC GCGGCGTACA ACTACTGTTT
 351 CGCGCGCGGC AGCGCATCTG CCAAACTACCA AATCAACGCG ATCCCGTTG
 401 CGACGCGCTT GGCGATACG GGCATGCGCA ACACGCGCGC CTATGAGCGC
 451 GTAGAAGTCG TGCGCGCGGT GCGCGGCGTG CTGGAACGGC CGGGCGAGCC
 501 TTTCCGCGCC GTCAATCTGG TGCGCAAAAG CCGCAGCGCG AAGCCATTGT
 551 TTGAAGTCCG CGCGCAAGCG GGCACACGCA AACATTTTCG GCTGGGCGCG
 601 GACGTATCGG GCAGCCTGAA TGCGCAAGGC ACGCTGCGCG GCCGCTGGT
 651 TTTCCACCTC GGACGCGCGC ACTCGTGGCG GCAGCGCGAA CGCAGCGCGG
 701 ATGCCGAACT CTACGCGATT TTGGAATACG ACATCGCACG GCACACCGCG
 751 GTCCACGCG AGCATGTGACTA CCAGCAGGCG AAGAAGACCG CGCAGCGCGC
 801 GCTCAGCTAC CGCGTGTACG ACAGCCCAAGG TTATCGCACC GCCTTCGGCC
 851 CGTSGCGGAA CCGCGCAACA AATTGCGGCA ACAGCGCGCA CGTGTGGCTG
 901 AACCTGTTTC CGCGCATGCA ACAGCGCTTC AACCAAGACT GGAAACTCAA
 951 AGCCGAATAC GACTACACCC CGACGCGCTT CGCGCAGCCG TACGCGTAT
 1001 CAGSGCTGCT TTCCATCGAG CACAACACCG CGCGCACCGA CTTGATTCCT
 1051 GGTATTGTCG ACGCGCAGCC GCGCACCCAC AGCGCCAGCG TGTCATTAA
 1101 CGGCAAAATAC CGCTGTGTCG GCGCGGAACA CGATTAAATC GCGGCTATCA
 1151 ACGGTTACAA ATACGCCAGC AACAAATACG GCGAACGACG CATCATCCCC
 1201 AACGCCATTG CCAACGCCCTA CGAATTTTTC CGCACGGGTG CTTACCGCGA
 1251 GCCTGCATCG TTTGCCCAAA CATCTCCGCA ATACGCGACC AGCGCGGAAA
 1301 TCGGCGGCTA TCTCGCCACC CGTTTCGCGC CGCGCGACAA CCTTTCGCTG
 1351 ATACTCGGCG GCAGATACAG CGGTTACGCG ACCCGCAGCT ACGACACGCG
 1401 CACACAAGGC ATGACCTATG TGTCGCGCAA CGSTTTTACC CCTACACAG
 1451 GCATCGTGTG CGACCTGACC GGCACCTGCT CGCTTACGCG CTGCTACAGC
 1501 AGCTGTTTCG TCCCGCAATC GCACAAGACG GAACACGGCA GCTACTCTAA
 1551 ACCGCTTACC GCGCAACAATC TCGAAGCGCG CATCAAGCGG GAAATGCTTT
 1601 AAGGCCGTCT GAACGCATCC GCCCGCGTGT ACCCGCGCGC TAAAAACAAC
 1651 CTCGCCACCG CAGCAGACCG GCACCCGAGC GGCACACACT ACTACGCGC
 1701 CGCCAAACCA GCCAAACACC ACGCGTGGGA AATCGAAGTC GCGCGCGCGA
 1751 TCACGCCCGA ATGCGAGATA CAGGCAGGTT ACAGCCAAAG CAAAACCCGC
 1801 GACCAGACCG GCAGCGGCTT GAACCCGACG AGCGTACCGC AAGCGAGCTT
 1851 CAAACTCTTC ACTCGCTTACC ACTTTGCCCC CGAAGCCCCC AGCGGCTGGA
 1901 CGATCGGCGC AGGCGTGCGC TGCGCAGAGC AAACCCACAC GCACCTTGCC
 1951 ACGCTCCGCA TCCCCAAACC CGCCGCCAAA GCGCGCGCGC CGCACAACAG
 2001 CGCGCAAAAA GCCTACGCGC TCGCGGCATC CATGGCGGCT TACGCTTCA
 2051 ATCCGCGCGC CGAAGCTGTC CTGAACGCTG ACAATCTGTT CAACAAACAC
 2101 TACCGCACCC AGCCCGACCG CCACGAGCTC GCGCGCACTG GCACAGTGAA
 2151 CGCGCGGTTT ACCTATGGGT TTAATAA

45 This encodes a protein having amino acid sequence <SEQ ID 668>:

1 MTRFKYSLLE AALLPVYAQA DVSVSDDEPK GESTELPTIT VTADRTASSN
 51 DGYTVSGTHT PIGLEMTIRE IPQSVSVITS QMRDONIKA LDRALLQATG
 101 TSRQIYGSDR AGYNYLFARQ SRIANYQING IPVADALADT GNANTAAYER
 151 VEVVRGVAGL LDGTGEPFAT VNLVRKRPTR KPLFEVRBAE GNRKHFGIGA
 201 DVSGSLNABG TLRGRILVSTF GRGDSWRQRE RSRDABLYGI LEYDIAPOTR
 251 VHAGMDYQQA KETADAPLSY AVYDSQGYAT AFGPKDNPAT NWANSRHRAL
 301 NLFAGIEHRF NQDWLKAERY DVTRSFRFRQ YGVGVLSDIT HNTAATDLIP
 351 GYWHADPRTH SASVSLIGKY RLFGRHDLI AGINGYKYAS NKYGERSIIP
 401 NAIPNAYEFS RTGAYPOQAS FAQTIPIQYGT RRIQIGVYAT RFRADNLSL
 451 ILGGYRSRYR TGSYDSRTQG MITYVSNRFT PYTGIVFDLT GNLISLYGSYS
 501 SLFVPSQSKD EHGSLKPVFT GNNLEAGIKG EWLEGRNLAS AAVYRARKNN
 551 LATAGRDES GNYVYRANQV AKTHGWEIV GCRITPEWQI QAGYSQKTR
 601 DQDGRSLNPD SVBERSFKLT TAYHFAPEAP SGWTIGAGVR WQSEHTHDEA
 651 TLRTPNPAK ARADNSRQK AYAVADIMAR YRNPRAELIS LNVNLFNRKH
 701 YRTQPDHRSY GALTVNAAF TYRFK*

ORF23a and ORF23-1 show 99.2% identity in 725 aa overlap:

orf23a.pep 10 20 30 40 50 60
 MTRFKYSLLEAALLPVYAQADVSDDEPKQUESTELPTITVTADRTASSNDGYTVSGTHT
 |||
 65 orf23-1 MTRFKYSLLEAALLPVYAQADVSDDEPKQUESTELPTITVTADRTASSNDGYTVSGTHT
 10 20 30 40 50 60

		70	80	90	100	110	120
	orf23a.pep	PLGLPMTLREIPQSVSVITSQQMRDQNIKALDRALLQATGTSRQIYGS DRAGYNLFAAG					
5	orf23-1	PLGLPMTLREIPQSVSVITSQQMRDQNIKALDRALLQATGTSRQIYGS DRAGYNLFAAG					
		70	80	90	100	110	120
	orf23a.pep	130	140	150	160	170	180
10	orf23-1	SRIANYQINGIPVADALADTGNANTAAAYERVEVVRGVAGLLDGTGEPSATVNLVRKFPTR					
		130	140	150	160	170	180
	orf23a.pep	190	200	210	220	230	240
15	orf23-1	KPLFEVRAEAGNRKHFGLGADVSGSLNAEGILRGLRVSTFGRGDSWRQRERSRDAELYGI					
		190	200	210	220	230	240
	orf23a.pep	250	260	270	280	290	300
20	orf23-1	LEYDIAFPQTRVHAGMDYQAKETADAPLSYAVYDSQGYATAFGPKDNPATNWSNRHRAL					
		250	260	270	280	290	300
25	orf23a.pep	310	320	330	340	350	360
	orf23-1	NLFAGIEHRRFNQDWKFLKAEYDYTRSFRQOPYGVAGVLSIDHNTAATDLIPGYWHADPRTH					
		310	320	330	340	350	360
30	orf23a.pep	370	380	390	400	410	420
	orf23-1	SASVSLGKYRLFGRHDLIAGINGYKYASKYGERSIIPNAIPNAYEFSRTGAYPOPAS					
35		370	380	390	400	410	420
	orf23a.pep	430	440	450	460	470	480
40	orf23-1	FAOTIPQYGTTRQIGGYLATRFRAADNLSLILGGRYRYRTGSDTSRTQGMTVVSANRPT					
		430	440	450	460	470	480
45	orf23a.pep	490	500	510	520	530	540
	orf23-1	PYTGI VFDLTGNLSYGSYSSLFVPSQSKDEHGSYLKPYTGNNLEAGIKGEWLEGRINAS					
		490	500	510	520	530	540
50	orf23a.pep	550	560	570	580	590	600
	orf23-1	AAVYRARKNNLATAAGRDPSGNTIYRAANQAKTHGWEIEVGGRIIPWQIQAGYSQSKTR					
		550	560	570	580	590	600
55	orf23a.pep	610	620	630	640	650	660
	orf23-1	DQDGSRLNPDSPVPSFKLFTAYHFAPEAPSGWTTGAGVRWQSETHTDPATLRINPAAK					
60		610	620	630	640	650	660
	orf23a.pep	670	680	690	700	710	720
	orf23-1	ARAADNSRQKAYAVADIMARYRFNPAELSLNVDNLFNKHRYTQPDHRSYGALRTVNAAF					
65		670	680	690	700	710	720
	orf23a.pep	TYRFXK					
70	orf23-1	TYRFXK					

Homology with a predicted ORF from *N. gonorrhoeae*

ORF23 shows 93.4% identity over a 211aa overlap with a predicted ORF (ORF23.ng) from *N.*

gonorrhoeae:

5	orf23.pep	GYNYLFARGSR IANYQINGIPVADALADTGNANTAAERVEVVRGVAGLLD	51
	orf23.ng	SAVDACRIPGYNLYLFARGSR IANYQINGIPVADALADTGNANTAAERVEVVRGVAGLLD	60
10	orf23.pep	GTGEP SATVNLVRKRLTRKPLFEVRAEAGNRKHFGLDADVSGSLNTEKKLRGLVSTFGR	111
	orf23.ng	GTGEP SATVNLVRKHPTKPLFEVRAEAGNRKHFGLDADVSGSLNAEGTLRGLVSTFGR	120
15	orf23.pep	GDSWRRRERSXAEYLGILEYDIAPOTRVHAXMDYQQAETADAPLSYAVYDSQGYATAF	171
	orf23.ng	GDSWRQLERSDAELYLGILEYDIAPOTRVHAGMDYQQAETADAPLSYAVYDSQGYATAF	180
	orf23.pep	GPKDNPATNWNASHRRALNLFAGIEHRFNQDWKLKAEYDY	211
	orf23.ng	GPKDNPATNWNSSNRNALNLFAGIEHRFNQDWKLKAEYDYTRSRFRQPYGVAGVLSIDHS	240

The ORF23.ng nucleotide sequence <SEQ ID 669> is predicted to encode a protein comprising

20 amino acid sequence <SEQ ID 670>:

1	SAVDACRIPG	YNYLFARGSR	IANYQINGIP	VADALADTGN	ANTAAYERVE
51	VVRGVAGLLD	GTGEP SATVN	LVRKHPTKRP	LFEVRAEAGN	RKHFGLDAD
101	SGSLNAEGLT	RGR LVSTFGR	GDSWRQLERS	DAELYLGILE	YDIAPOTRVH
151	AGMDYQQAKE	TADAPLSYAV	YDSQGYATAF	GPKDNPATNW	NSNRNALNLT
201	FAGIEHRFNQ	DWKLKAEYDY	TRSRFRQPYG	VAGVLSIDHS	TAATDLIPGY
251	WHADPRTHSA	SMSTLTKYRL	FGREHDLIAG	INGKYASANK	YGRSIIIPNA
301	IPNAYEFERT	GAYPQSSSFA	QTIPQYDTRR	QIGGYLATRF	RAADNLSLIL
351	GRHYRRRAAG	YVNSYTGQMT	TVSRAETFPY	GIIVEDLTGN	LSLIGYSYSL
401	FYPQLQKDEH	GSYLLFPVTG	NLEADIKGEH	LEGRINASLA	VYRARKNWLA
451	TAAGRDQSGN	TYRAANQAK	THGWIEVGG	RITPEWQTQA	GYSQSPKRDQ
501	DGSRINPDV	PERSEKFLTA	YHLAPEAFSG	RTIAGVRRQ	GETHTDPAAL
551	RIPNPAKAKR	AVANSRQKAY	AVADIMARYR	FNRTLESLN	VDNLFNKHRY
601	TQDPDRHSYGA	LRTVNAFTY	RFK*		

Further work revealed the complete nucleotide sequence <SEQ ID 671>:

35	1	ATGACACGCT	TCAAATACTC	CCTGCTTTTT	GCGGCCCTGC	TACCGGTGTA
	51	CGCGCAGGCC	GATGTTCTCG	TTTCAGACGA	CCCCAACCAG	CAGGAAAGCA
	101	CGGAATTGCC	GACCATCAC	GTTACCGCGC	ACCGCACCGC	GAGTCCACAC
	151	GACGGCTACA	CGTTTCCGG	CACGACACCC	CGGTCGGGCG	TGCCCATGAC
	201	CCTGGCGGAA	ATCCCGCAGA	GCGTCAGCGT	CATCACATCG	CAACAAATGC
40	251	GCGACCAAAA	CATCAAAACG	CTCGACCGCG	CCCTGTTGCA	GCGACCGCG
	301	ACGACCGCGC	AGATTATCGG	CTCGACCGCG	CGCGGCTACTA	ACTACCTGTT
	351	CGCGCGCGC	ACGCGATCG	CCACCTACGA	ATCCACCGCG	ATCCGCTGCG
	401	CGACCGCGCT	GGCGGATACG	GGCAATGCCA	ACACCGCGCG	CTATTGAGCG
	451	GTAGAAGTGG	TGCGCGCGCT	GGCGGGGCTG	CCGACCGGCA	CGGCGGAGCC
45	501	TTCTGCCACC	GTCAATCTGG	TACGCAACAA	CCGACCGCGC	AAGCAATTGT
	551	TTGAAGTCCG	CGCGGAAGCC	GCGACCGGCA	AACATTTCGG	GCTGGGCGCG
	601	GACGTATCGG	GCAAGCTGAA	CGCGGAAGCG	ACGCTGCGCG	GCGGCTGGGT
	651	TTCCACCTTC	GGAAGCGGGG	ACTCGTGGCG	GCAAGCTGAA	CGCAGCGCGG
	701	ATGCGCGAAT	CTACGGCAAT	TTGGAATACG	ACATCGCACC	GCAACCCGCG
50	751	GTCACGCGAG	GCAATGAGCTA	CCAGCAGCGC	AAAGAAACCG	CAGACCGGCC
	801	GCTCAGCTAC	CGCGGTGACG	ACAGGCCAAG	TTATGCCACC	CGCTTCGGCG
	851	CAAAAGACAA	CCCGGCCACA	AATTGGTCTGA	ACAGCGCGAA	CGTGCGGCTC
	901	AACCTGTTGG	CGCGCATAGA	ACACCGCTTC	AACCAAGACT	GGAAACTCAA
	951	AGCGGATATC	GACATACACC	GTAGCGCTTC	CGCCAGGCC	TACGGGTGGG
55	1001	CAGCGTACT	TTCCATCGAC	CACAGCAGCT	CGCCACCGA	CTCGATTCCC
	1051	GGTATTGGCG	ACGCGagatcc	GCGCACCCAC	AGCGCGACGA	TGTCATTGAC
	1101	CGGCAAAATAC	CgcctGTTCC	GCGCGGAGCA	CGATTAAATC	GCGGATATCA
	1151	ACGGCTACAA	ATACGCCGAGC	AACAATATAG	GCGAAGCGAG	CATCATTTCC
	1201	AAGCGCAATC	CCAGAGGCTA	CGAATTTTCC	CGCACGGGGG	CTTATCGCGA
	1251	GCCATCATCG	TTTGCCCAAA	CCATCCCGCA	ATACGACACC	AGGCGGCAAA
60	1301	TCGGGCGGCTA	TCTGCCCAAC	CGTTTCCGCG	CGCGCGACAA	CTTTGCGCTG
	1351	ATATCTCGGG	GCAGATACAG	CGCGTACCGC	GCAGGACAGT	ACAACAGCGG

1401	CACACAAGGC	ATGACCTATG	TGTCGGCCAA	CGGTTTCACC	CCCTACACAG
1451	GCATCGTGTT	CGATCTGACC	GGCAACCTGT	CGCTTTACGG	CTCGTACACG
1501	AGCCTGTTCG	TCCGCAATT	GCAAAAAGAC	GAACACGGCA	GCTACCTGAA
1551	ACCCGTAAAC	GGCAACAATC	TGSAAGCCGA	CATCAAAAGC	GAATGCGTTG
1601	AAGGGCGTCT	GAACGACATC	GCCCGCTGT	ACCGGCGCTG	TAAAAACAC
1651	CTCGCCACCT	CACAGAGAG	CGACAGAGC	GGCAACAGCT	ACTATCGCC
1701	CGCCAGCCAA	GCACAAACCC	ACGCGTGGGA	AATCGAAGTG	GGCGGCCGCA
1751	TCACGCCCGA	ATGCGAGATA	CAGGCAGGCT	ACAGCCAAAG	CAACCCGCC
1801	GACCAAGAGC	CGACGCGCTC	GAACCCGAC	ACGCTACCG	AACGAGCTT
1851	CAACTCTTTC	ACCGCCTACC	ACTTAGCCCC	CGAAGCCCCC	AGCGGCCGGA
1901	CCATcggTGC	GGGTGTGCG	CGCGAGGGCG	AAACCCACAC	CGACCCAGCC
1951	GGCGTCCGCA	TCCCAACCC	CGCCGCGCAA	GCCCGGCGCG	TCGCCAACAG
2001	CGCGCAGAAA	GCTACGCGC	TCGCGGCACAT	CATGGCGCGT	TACCGCTTCA
2051	ATCCGCGCAC	CGAACTGTGC	CTGAACGTGG	ACAACTGTGT	CAACAAACAC
2101	TACCGCGACC	AGCCGACCC	CCACAGCTAC	GGCGCACTGC	GGACAGTGAA
2151	CGCGCGGTTT	ACCTATCGGT	TTAAATAA		

This corresponds to the amino acid sequence <SEQ ID 672; ORF23ng-1>:

1	MTRFKYSLLF	AALLPVYAQA	DVSVSDPKP	QESTELPTIT	VTADRTASSN
51	DGYTSGTHT	PFGLMTLRE	IPQSVSVITS	QOMRDQNIKT	LDRLALQRTG
101	TSRQYIGSDR	AGYNLFARG	SRIANYQING	IPVADALADT	GNANTAYER
151	VEVVRGVAGL	PDGTGFSAT	VNLVRKHPT	KPLFEVRAEA	GNRKHFGLCA
201	DVSGSLNAES	TLRGLRVSTF	GRGDSWRQLE	RSRDAELYGI	LEYDIAPOTR
251	VHAGMDYQQA	KETADAPLSY	AVYDSQGYAT	AFGPKDNFAT	NWSNRNRAL
301	NLFAGIEHRF	NDQWKLKAEY	DYTRSFRFQP	YGVAGVLSID	HSTAATDLIP
351	GYWHADPRTH	SASMSLTGKY	RLEFGREHDLI	AGINGYKYAS	NKYGERSIIP
401	NAIPNAYEFS	RTGAYPOPS	FAQTIPOYDT	RROQIGYLAT	RFRADNLSL
451	ILGGYRSRYR	AGSYNSRTQG	MTYVSNRFT	PYTGIVFDLT	GNLSLYGSYS
501	SLFVPQLQKD	EHGSYLKPTV	GNNLEADIKG	EWLEGRNLAS	AAVYRARKNN
551	LATAAGRDQS	GNTYYRAANO	AKTHWEIEIV	GGRTIPEWQI	QAGYSQSKFR
601	QDQGSRLNPD	SVERSEFKLF	TAYHLAPEAF	SGRTIGAGVR	RGETHITDEA
651	ALRIPNPAAR	ARAVANSRQK	ATAVADIMAR	YRNRPTLETS	LNVDNLFNKH
701	YRTQDRHSY	GALRTVNAAF	TYRFK*		

ORF23ng-1 and ORF23-1 show 95.9% identity in 725 aa overlap:

		10	20	30	40	50	60
35	orf23-1.pep	MTRFKYSLLFAALLPVYAQA	DVSVSDPKPQESTELPTIT	VTADRTASSN	MDGYTVSGTHT		
	orf23ng-1	MTRFKYSLLFAALLPVYAQA	DVSVSDPKPQESTELPTIT	VTADRTASSN	MDGYTVSGTHT		
		10	20	30	40	50	60
40	orf23-1.pep	PLGLFMTLREIPQSVSVITS	QOMRDQNIKT	LDRLALLQATGTSRQYIGSDR	AGYNLFARG		
	orf23ng-1	PLGLFMTLREIPQSVSVITS	QOMRDQNIKT	LDRLALLQATGTSRQYIGSDR	AGYNLFARG		
		70	80	90	100	110	120
45	orf23-1.pep	SRIANYQINGIPVADALADT	GNANTAYERVEVVRGVAGL	PDGTGFSATVNLVRKHPT			
	orf23ng-1	SRIANYQINGIPVADALADT	GNANTAYERVEVVRGVAGL	PDGTGFSATVNLVRKHPT			
		130	140	150	160	170	180
50	orf23-1.pep	KPLFEVRAEAGNRKHFGLD	ADVSGSLNTEGTLRGLRVSTF	GRGDSWRQLERSDAELYGI			
	orf23ng-1	KPLFEVRAEAGNRKHFGLD	ADVSGSLNTEGTLRGLRVSTF	GRGDSWRQLERSDAELYGI			
		190	200	210	220	230	240
55	orf23-1.pep	LEYDIAPOTRVHAGMDYQQA	KETADAPLSYAVYDSQGYAT	AFGPKDNFATNWSNRNRAL			
	orf23ng-1	LEYDIAPOTRVHAGMDYQQA	KETADAPLSYAVYDSQGYAT	AFGPKDNFATNWSNRNRAL			
		250	260	270	280	290	300
60	orf23-1.pep	NLFAGIEHRFNQDWKLKAEY	DYTRSFRFQPGYVAGVLSID	HNTAATDLIPGYWHADPRTH			
	orf23ng-1	NLFAGIEHRFNQDWKLKAEY	DYTRSFRFQPGYVAGVLSID	HNTAATDLIPGYWHADPRTH			
		310	320	330	340	350	360
65	orf23-1.pep	NLFAGIEHRFNQDWKLKAEY	DYTRSFRFQPGYVAGVLSID	HNTAATDLIPGYWHADPRTH			
	orf23ng-1	NLFAGIEHRFNQDWKLKAEY	DYTRSFRFQPGYVAGVLSID	HNTAATDLIPGYWHADPRTH			

-385-

		310	320	330	340	350	360
		370	380	390	400	410	420
5	orf23-1.pep	SASVSLIGKYLFGREHDLIAGINGKYKASNKYGRSIIIPNAIPNAYEFSRTGAYPOFAS					
	orf23ng-1	SASMSLTGKYLFGREHDLIAGINGKYKASNKYGRSIIIPNAIPNAYEFSRTGAYPOFAS					
		370	380	390	400	410	420
10	orf23-1.pep	FAQTIPOYGTTRRQIGGYLATRFRAADNLSLILGGRYTRYRTGSYDSRTQGMTYVSANRFT					
	orf23ng-1	FAQTIPOYDTRRQIGGYLATRFRAADNLSLILGGRYTRYRTGSYDSRTQGMTYVSANRFT					
		430	440	450	460	470	480
15	orf23-1.pep	PYTGIYFDLTGNLSLYGSYSLSFVPSQKDEHGSYLPKFTGNLLEAGIKGEWLEGRNLAS					
	orf23ng-1	PYTGIYFDLTGNLSLYGSYSLSFVPSQKDEHGSYLPKFTGNLLEAGIKGEWLEGRNLAS					
		490	500	510	520	530	540
20	orf23-1.pep	AAVYRARKNNLATAAGRDPGSGNTYYRAANQAKTHGWEIEVGGRIITPEWQIQAGYSQSKTR					
	orf23ng-1	AAVYRARKNNLATAAGRDPGSGNTYYRAANQAKTHGWEIEVGGRIITPEWQIQAGYSQSKTR					
		550	560	570	580	590	600
25	orf23-1.pep	DQDGSRLNPDSVPERSFKLFTAYHFAPEAPSQWTIGAGVVRQSEHTHPDPAALRINPAK					
	orf23ng-1	DQDGSRLNPDSVPERSFKLFTAYHFAPEAPSQWTIGAGVVRQSEHTHPDPAALRINPAK					
		610	620	630	640	650	660
30	orf23-1.pep	ARAADNSRQKAYAVADIMARYRFNPRARLSLNVNLFNKHYRTQPDHRYGALRTVNAAF					
	orf23ng-1	ARAADNSRQKAYAVADIMARYRFNPRARLSLNVNLFNKHYRTQPDHRYGALRTVNAAF					
		670	680	690	700	710	720
35	orf23-1.pep	ARAADNSRQKAYAVADIMARYRFNPRARLSLNVNLFNKHYRTQPDHRYGALRTVNAAF					
	orf23ng-1	ARAADNSRQKAYAVADIMARYRFNPRARLSLNVNLFNKHYRTQPDHRYGALRTVNAAF					
		670	680	690	700	710	720
40	orf23-1.pep	TYRFXK					
	orf23ng-1	TYRFXK					

In addition, ORF23ng-1 shows significant homology with an OMP from *E. coli*:

45	splP16869 FHUE_ECOLI OUTER-MEMBRANE RECEPTOR FOR FE(III)-COPROGEN, FE(III)-FERRIOXAMINE B AND FE(III)-RHODOTRULIC ACID PRECURSOR >gi11651542 gnl PID d1015403 D90745 Outer membrane protein FhuE precursor [Escherichia coli] >gi11651545 gnl PID d1015405 D90746 Outer membrane protein FhuE precursor [Escherichia coli] >gi11787344 AE000210 outer-membrane receptor for Fe(III)-coprogen, Fe(III)-ferrioxamine B and Fe(III)-rhodotrulic acid precursor [Escherichia coli] Length = 729
50	Score = 332 bits (843), Expect = 3e-90 Identities = 228/717 (31%), Positives = 350/717 (48%), Gaps = 60/717 (8%)
55	Query: 38 TITVTADRTASSN--DGYTVSGTHTFPGFLMFLREIPQSVSVITSSQMDRDNKTLDRAL 95 T+ V TA + + Y+V+ T + MT R+IPQSV+++ Q-M DQ +TL+ + Sbjct: 43 TVIVEGSAATPDDGENDYSVTSAGTRMQMTQRDIPQSVTVISQQRMDQQQLTLGEMV 102
60	Query: 96 LQAGETSRQIYSGSDRAGYNILFARGSRITANYQINGIP-----VADALADTGNANTAA 147 G S+ SDRY + +RG +I NY + +GIP + DAL+D A Sbjct: 103 ENTLSISKSQADSRLALY---YSGRFQIDNYMVDGIPTFYFESWNLGDALSDM-----AL 154
65	Query: 148 YERVEVVRGVAGLFDGTGEPSATVNLVRKHPTKFLF-EVRAEAGNRKHIFGLGADVSGSL 206 +ERVEVVRG GL GTG PSA -N+VRKH T + +V AE G+ AD+ L Sbjct: 155 FERVEVVRGATGLMTGTGNPSAALNVRKHATSRFPGKDVSAEYGSWKNERYVADLQSL 214
70	Query: 207 NAEGLTGRGLVSTFGRGDSWRQLERSRDAELYGILEYDIAPOTRVHAGMDYQAKETADA 266 +G+R R+V + DSW S GI++ D- T + AG +YQ+ + Sbjct: 215 TEDGKIRARIVGGYQNNDSWLDTRYNSEKTFEFGIVDADLGLDGLTLAGYEQRIQDVNSPT 274
	Query: 267 PLSYAYVDSQGYATAFGPKDNFATNWSNRRLNLFAGIEHRFNQWKLKAEYDYDTRSR 326

-386-

+++ G + ++ + A +W+ + +F + +F W+ ++
 5 Sbjct: 275 WGGLPRWNTDGSNSYDRARSTAPDWAYNDKEINKVEMTLKQQFADTWGATLNATHSEVE 334
 Query: 327 F--RQPYGVAGVLSIDHSTAA--TDLIPGY-----WHADPHTSA-SMSLTGKYRLFG 374
 F + Y A V D ++ PG+ W++ R A + G Y LFG
 Sbjct: 335 FDSKMMYVDAYVKNADGMLVGFPYSNYGGFDYVGGTGWNSGKRVKDALDLFADGSEYELFG 394
 10 Query: 375 REHDLIAGINGKYVASNKYGER--SIIPNAIPNAVEFSRTGAYPQSSFACTIPQYDTRR 432
 R+H+L+ G Y +N+Y +I P+ I + Y F+ G +PQ Q++ Q TD
 Sbjct: 395 RQHNLMPG-GSYSKQNNRYFSSWANTFPDEIGSFYFN--GNFPQTDWSPSPSLAQDQDTH 451
 15 Query: 433 QCGGLATRFRAADNLNLILGGGRYSRYRAGSYNSRTQGMTY-VSANRFTPTTYGIYFDXX 491
 Y ATR AD L LILG RY+ +R + +TY+ N TPY G+VFD
 Sbjct: 452 MKSLYAATRTVLADPLHLILGARYTNWRVDT-----LTSMEKNHTTPYAGLVFDIND 504
 20 Query: 492 XXXXXXXXXXXXVFPQLQKDEHGSYLKFPVTGNNLEADIKGEWLEGRINASAAYVYARKNNL 551
 F PQ +D G YL P+TGNN E +K +W+ RL + A++R ++N+
 Sbjct: 505 NWSTYASYTSIFQFNDRDSSGKYLAFITGNNYELGLKSDWMNSRLTTTLAIFRIEQDNV 564
 25 Query: 552 ATAAGR---DQSGNTYYRAANQAKTHGWEIEVGGRIITPEWQIQAGYSQSKPRQDQGSRLN 608
 A + G +G T Y+A + + G E E+ G IT WQ+ G ++ D +G+ +N
 Sbjct: 565 AGSTGTGPIPGSNGETAYKAVDGTVSKGVFELNGAITDNWQLTFGATRYIAEDNEGNAYN 624
 30 Query: 609 PDSVPERSFKLFTAYHLAEPASGRTIGAGVRRQGETHTDPAALRIPNPAAKARAVANSR 668
 P ++P + K+PT+Y L P P T+G GV Q +TD P RA
 Sbjct: 625 P-NLPTTVMKFTSYRL-FVMPE-LTVGGGVNWQNRVYDTV-----TPYGTFR-----F 672
 Query: 669 QKAYAVADIMARYRNPRTSLNVDNLNKHRYTQPDHR-SYGLALRTVNAATYFR 724
 Q +YA+ D+ RY+ L NV+NLP-K Y T + YG R + TY+F
 Sbjct: 673 QSGYALVDLFTYQVTKNFSLQGNVNNLFDKTYTDNVEGSIYVGPFRNRSITGTQGF 729

Based on this analysis, it was predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF23-1 (77.5kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 15A shows the results of affinity purification of the His-fusion protein, and Figure 15B shows the results of expression of the GST-fusion in *E.coli*. Purified His-fusion protein was used to immunise mice, whose sera were used for Western blot (Figure 15C) and for ELISA (positive result). These experiments confirm that ORF23-1 is a surface-exposed protein, and that it is a useful immunogen.

Example 80

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 673>:

1 ATGCGCACGG CAGTGGTTT GCTGTGATC ATGCCGATGG CGGCTTCGTC
 51 GGCAATGATG CCGGAATGCG GTGTGCGGGG CGTGTGCGCG GGAACGGCAA
 101 TCATATCCAA GCCGACCGAA CAACGCGCGG TCATGGCTTC GAGTTTGTCC
 151 AGCGTCAGcA CGCTTGCTTC GCGCGcGgCa ATCATCACTT CGTCTTCGGA
 201 AACGGGGATA AACGC-GCCAC TCACCAACCCG GACCGCGCTG GAAGCCATCA
 251 TGCCGCCCTTT TTTCAAGGCA TCGCTTCAGCA ATGCCAAGC TGCTGTGTGTG
 301 CCGTGCCTAC CGCAGACGCT CAAGCCCATTT ThTTCAGAA TGCGTGCCAC
 351 ThAGTCGCCG ACGGGG..

This corresponds to the amino acid sequence <SEQ ID 674; ORF24>:

1 MRTAVVLLLI MFAWASSAMH FEMVCAGVSP GTALISKFTF QTAVMASSLS
 51 SVSTPASAAA IIPSSSETGI NAPLKPPAL EAIMPPFFTA SFSNAKAAYV
 101 FCVPEQLKFI XSSMRATXSP TG..

Further work revealed the complete nucleotide sequence <SEQ ID 675>:

```

1  ATGGCGCAGG CAGTGGTTTT GCTGTTGATC ATGCCGATGG CGGCTTCGTC
51  GGCAATGATG CCGGAATAGG TGTGGCGGG CGTGTGCGCG GGAACGGCAA
101 TCATATCCAA GCCGACGGAA CAACCGGGG TCATGGCTTC GAGTTTGTCC
151 AGCGTCAGCA CGCTGCGTTC GGGCGGGGCA ATCATGCTTC CGTCTTCGGA
201 AACGGGGATA AACGGCCAC TCAGACCGCC GACCGGGTGT GAAGCCATCA
251 TGCGGCGCTT TTTCAGCGCA TCGTTCAGCA ATGCCAAGC TGCTGTTGTG
301 CCGTGGGTAC CGCAGACGCT CAAGCCCAT TCTTCAAGAA TGGGTGCCAC
351 TGAGTGGCGC ACGGCGGGGG TCGGGGCCAG CGACAAGTGT AGAATACCAA
401 ACGGGATATT CAGCATTTTT GAGGCTTCGC GCGCGATGAG TTCGCCACG
451 CGGGAATATT TGAAAGCAGT TTCTTCACCT ACTTCGCAA CTTCGGTCAA
501 TGTGTTGCA TCTGAATTT CCAACCGGC TTTCAGCA CCGTGGCGCG
551 ATACGCGCAC ATTGATACG GCATCGGCTT CGCCGCAACC ATGAACCGCG
601 CCGGCCATAA ACGGTTTGTG TTCCACCGCG TTGCGAACA CGACAATTTT
651 AGCGCAGCGC AAGCCTTCGG CGGTGATTTC CGCCGTGGT TTGACGGTTT
701 CGCCCGCCAG CTTGACCGCA TCCATATTGA TACCGCGACG CGTACTGCCG
751 ATATTGATGG AGCTGCACAC AATATCGGTA GTCTTCATCG CTTCGGGAAT
801 GGAGCGGATT AACACCTCAT CCGAAGCGCA CATCCCTTTT TGCACCAACG
851 CGGAAAAACC GCCGATAAAA GACACACCGA TGGCTTTGCG AGCTTTATCC
201 901 AAAGTTTGGC CCACGCTGAC GTAA

```

This corresponds to the amino acid sequence <SEQ ID 676; ORF24-1>:

```

1  MRTAVVLLLI MPMAASSAMM PEMVCAGVSP GTAIISKPTE QTAVMASSLS
51  SVSTPASAAA IIPSSSETGI NAPLKPPTAL EAIMPFPEFTA SFNNAKAHV
101 PCVPQTLKPI SRKMRATESP TAGVGASDKS RIFNGIFSTF EASRPMSEPT
151 RVILKAVFTT TSATSVNVA SEFSNAATT PGEPPTLIT ASASEP+NA
201 PAINGLSSTA LQNTTILAQP KPSGVISAVR LTVSPASLTA SILIPARVLE
251 ILMELHTISV VFIAAGMERI NTSSEGDTPF CTNAEKKPKK DTPMALAALS
301 KVCAITLT*

```

Computer analysis of the amino acid sequence gave the following results:

30 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF24 shows 96.4% identity over a 307 aa overlap with an ORF (ORF24A) from strain A of *N. meningitidis*:

	10	20	30	40	50	60
35 orf24a.pep	MRTAVVLLLI	MPMAASSAMM	PEMVCAGVSP	GTAIIISKPTE	QTAVMASSLS	SNVSTPASAAA
orf24						
	MRTAVVLLLI	MPMAASSAMM	PEMVCAGVSP	GTAIIISKPTE	QTAVMASSLS	SNVSTPASAAA
	10	20	30	40	50	60
40 orf24a.pep	IIPSSSXTG	INAPLKPPTALE	AIMPPFFTA	SFNSNAKAHV	PCVPQTLKPI	SSRMRATESP
orf24						
	IIPSSSXTG	INAPLKPPTALE	AIMPPFFTA	SFNSNAKAHV	PCVPQTLKPI	SSRMRATESP
	70	80	90	100	110	120
45 orf24a.pep	TAGVGASDKS	RIPNGIFSI	FEASRPMSS	PTRVILKAVFTT	SATSVNVA	SEFSNAATT
orf24						
	TAGVGASDKS	RIPNGIFSI	FEASRPMSS	PTRVILKAVFTT	SATSVNVA	SEFSNAATT
	130	140	150	160	170	180
50 orf24a.pep	PGFDPPTLIT	ASASEPFXNA	PAIXGLSSKAL	QNTTILAQPKP	SSVISXVRLM	VMSPASLTA
orf24						
	PGFDPPTLIT	ASASEPFXNA	PAIXGLSSKAL	QNTTILAQPKP	SSVISXVRLM	VMSPASLTA
	190	200	210	220	230	240
55 orf24a.pep	SILIPARVLE	ILMELHTISV	VFIAAGMERI	NTSSEGDTPF	CTNAEKKPKK	IDTPMALAALS
orf24						
	SILIPARVLE	ILMELHTISV	VFIAAGMERI	NTSSEGDTPF	CTNAEKKPKK	IDTPMALAALS
	250	260	270	280	290	300
60 orf24	SILIPARVLE	ILMELHTISV	VFIAAGMERI	NTSSEGDTPF	CTNAEKKPKK	IDTPMALAALS
	250	260	270	280	290	300

-388-

```

orf24a.pep  KVCATLTX
            |||||
orf24       KVCATLTX

```

- 5 The complete length ORF24a nucleotide sequence <SEQ ID 677> is:

```

1  ATCGGCACGG  CAGTGGTTTT  GCTGTTGATC  ATGCCGATGG  CGGCTCGTC
51  GCGAATGATG  CCGGAAATGG  TGTGCGCGGG  TGTGTCGCCG  GGAACGGCAA
101  TCATATCCAA  NCCGACCGAA  CAAACGCGGG  TCATCGCTTC  GAGTTTATCC
151  AACGTACGCA  CGCTGCTTTC  GGGCGGGGCA  ATCATACCTT  CGTCTCGGCA
201  NACGGGGATA  AAGCGGCCAC  TCAACCGGCC  AACCGCGCTC  GAAGCATCA
251  TGGCGCCCTT  TTTCACGGCA  TCGTTCAGCA  ATGCCAAGCG  TGCTGTTGTG
301  CGGTGGGTAC  CGCAGACGGT  CAAACCCATT  TCTTCAAGAA  TGGCGGCCAC
351  CGAGTCGCCG  ACGGCAGGGG  TCGGTGCCAG  CGACAAGTCG  AGAATACCAA
401  ACGGGATATT  CAGCATTTTT  GAGGCTTCGC  GGCGGATGAG  TTGCGCCACG
15  451  CGGGTAATTT  TGAAGGCGGT  TTCTTTCACA  ACTTGGCCAA  CTTCGCTCAA
501  TGTGCTTGCA  TCGAATTTT  CCAACGCGGG  TTTTACGACA  CCGGGGCGGG
551  ATACGCGGAC  ATTAATCACA  GCATCGCGTT  CGCTGAGCC  GTGAAACGGG
601  CCGGCCATAN  ACGGGTTGTC  TTCCNCGCG  TTGCAGACGA  CGACGATTTT
651  GCGCGACGCG  AAGCCTCTTA  GTGTGATTTC  ACGCTGCGT  TTGATGCTTT
701  CGCCCGCCAG  TCTGACCGGG  TCCATATTGA  TACCGCGCGG  CGTACTGCGG
751  ATATTGATGG  AGCTGCACAC  GATATCAGTA  GTCTTCATCG  CTTCGGGAAT
801  GGAACGGGAT  AACACCTCGT  CAGAAGGCGA  CATACCTTTT  TGCACGACGG
851  CGGAAAGGCC  GCCAATAAAA  GACACGCCGA  TGGCTTTGGC  AGCCTTATCC
901  AAGTTTGGC  CCACGCTGAC  GTAA

```

- 25 This encodes a protein having amino acid sequence <SEQ ID 678>:

```

1  MRTAVVLLLI  MFMAASSAMM  PEMVCAGVSP  GTAIISXPTE  QTAVIASSLS
51  NVSTPASAAA  IIPSSXGTGI  NAPLKPPTAL  EAIMPPFFTA  SFSNAKAADV
101  PCVPQTLKPI  SSRMRATESP  TAGVGADSKS  RPNPFIPIFI  EASRPMSPST
151  RVILKAVFFT  TSATSVNVVA  SFSNNAAFIT  PGPDPTTLIT  ASASPEP*NA
201  PAIXGLSSXA  LQNTTILAQP  KPSSVISXVR  LMSVPASLTA  SILIPARVLP
251  IIMELHTISV  VFASGMERX  NTSSEGDIFP  CTSAEKFPK  DFMALAALS
301  KVCATLT*

```

It should be noted that this protein includes a stop codon at position 198.

ORF24a and ORF24-1 show 96.4% identity in 307 aa overlap:

```

35  orf24a.pep      10      20      30      40      50      60
      MRTAVVLLIIMPMAASSAMMPPEMVCGVSPGTAIISXPTEQTAVIASSLSNVSTPASAAA
      orf24-1      10      20      30      40      50      60
      MRTAVVLLIIMPMAASSAMMPPEMVCGVSPGTAIISXPTEQTAVIASSLSNVSTPASAAA

40  orf24a.pep      70      80      90      100     110     120
      IIPSSXGTGINAPLKPPTALEAIMPPFFTASFSNAKAADVPCVPQTLKPISSRMRATESP
      orf24-1      70      80      90      100     110     120
      IIPSSXGTGINAPLKPPTALEAIMPPFFTASFSNAKAADVPCVPQTLKPISSRMRATESP

45  orf24a.pep      130     140     150     160     170     180
      TAGVGADSKSRPNPFIPISEASRPMSSPTRVILKAVFFTTSATSVNVVASEFSNNAFTT
      orf24-1      130     140     150     160     170     180
      TAGVGADSKSRPNPFIPISEASRPMSSPTRVILKAVFFTTSATSVNVVASEFSNNAFTT

50  orf24a.pep      190     200     210     220     230     240
      PGPDPTTLITASASPEPKNAPAIIXGLSSXALQNTTILAQPKPSSVISXVRLMVPASLTA
      orf24-1      190     200     210     220     230     240
      PGPDPTTLITASASPEPKNAPAIIXGLSSXALQNTTILAQPKPSSVISXVRLMVPASLTA

55  orf24a.pep      250     260     270     280     290     300
      SILIPARVLPILMELHTISVVFASGMERNXTSSSEGDIFPCTSAEKFPKIDTFMALAALS
      orf24-1      250     260     270     280     290     300
      SILIPARVLPILMELHTISVVFASGMERNXTSSSEGDIFPCTSAEKFPKIDTFMALAALS

```

orf24a.pep KVCATLTX
 5 orf24-1 KVCATLTX

Homology with a predicted ORF from *N.gonorrhoeae*

ORF24 shows 96.7% identity over a 121 aa overlap with a predicted ORF (ORF24ng) from

N.gonorrhoeae:

10	orf24.pep	MRTAVVLLILIMPAASSAMPEMVCAGVSPGTAIISKPTTEQTAVMASSLSVSTPASAAA	60
	orf24ng	MRTAVVLLILIMPAASSAMPEMVCAGVSPGTAIMSKPTEQTAVMASSLSVNTPASAAA	60
15	orf24.pep	IIPSSSETGINAPLKPPTALEAIMPPFFTASFNSAKAAVVPVQTLKPIKSRMRATXSP	120
	orf24ng	IIPSSSETGINAPLKPPTALEAIMPPFFTASFNSAKAAVVPVQTLKPIISSRMRAATESP	120
	orf24.pep	TG	122
	orf24ng	TAGVGASDKSRMPNGIFSIFFASRPMSSPTRVILKAVFFTTTSATSVRLTASEFSSAALTT	180

The complete length ORF24ng nucleotide sequence <SEQ ID 679> is:

1	ATGCCGACGG	CGGTGGTTTT	GCTGTTGATC	ATGCCGATGG	CGGCTTCGTC
51	GGCGATGATG	COGGAATG	TGTGCGCGGG	CGTGTGCGCG	GGAAACGGCAA
101	TCATGTCCAA	ACCAACGGAG	CAGACGGGGG	TCATGGCTTC	GAGTTTGTCC
151	AGCGTCAACA	CGCTGCGCTC	GGCGGGCGCA	ATCATACCTT	CGTCTTCGGA
201	AACGGGGATA	AACCGCGCGC	TCAACCGCCG	GACCGCGCTG	GAGCGCATCA
251	TGCGCCCTTT	TTTCAAGGCA	TGCTTCAGCA	ATGCCAAGC	TGCTGTGTG
301	CGCTGCGTAC	CGCAGACGCT	CAGCCCATTT	TCCTCAAGAA	TGCGCGGCAC
351	CGATGCGCGC	ACGCGGGGGG	TGCTGCGCAG	CGACAAATCG	AGAATCGCA
401	ACGGGATATT	CAGCATTTTT	GAGGCTTCGC	GACCGATGAG	TCGCCACAG
451	CGGGTGATTT	TGAAAGCGGT	TTTCTTCACG	ACTTCGCGCA	CCTCGCTCAG
501	GCTGACCGCG	TCCGAATTTT	CGAGCGCGCG	TTTGACCAAC	CCTGGACCGG
551	ATACGCCGAC	ATTAAATCACA	GCATCCGGCT	CGCCCGGACC	GTGGAACGCA
601	CCCGCCATAA	ACGGATTGTC	TTCCACCGCG	TTGCAGAAC	CGACGATTTT
651	GGCGCAGCGC	AAACCTTCGG	GTGTGATTTT	AGCCGTGCGT	TTGATGGTTT
701	CGCTTGCCAG	CTTGACCGCA	TCCATATTGA	TACCGGCACG	CGTGTGCGCG
751	ATATTGATGG	AGCTGCACAC	GATATCGGTA	GTTTTCATCG	CTTCGGGAAC
801	GGAAACGGATC	AACACCTCAT	CCGAAGCGCA	CATACCTTTT	TGCACGACGC
851	CGGAAAGACC	GCCGATAAAG	GACACGCCGA	TGGCTTTGGC	TGCGTTGTTC
901	AAAGTCTGCG	CCAGCGTGAC	ATAA		

This encodes a protein having amino acid sequence <SEQ ID 680>:

1	MRTAVVLLI	MPMAASSAM	PEMVCAGVSP	GTAIMSKPTE	QTAVMASSLS
51	SVNTPASAAA	IIPSSSETGI	NAPLKPPTAL	EAIMPPFFTA	SFSNAKAAV
101	PCVPQTLKPI	SSRMRAATESP	TAGVGASDKS	RMNGIFSIFF	EASRPMSSPT
151	RVILKAVFFT	TSATSVRLTA	SEFSSAALTT	PGFDPTTLIT	ASASPEPNA
201	PAINGLSSTA	LQNTTILAQP	KPSGVISAVR	LMVSPASLTA	SILIPARVLP
251	IIMELHTISV	VFIASGTERI	NTSSEGDIFP	CTSAEKFFIK	DTPMALALS
301	KVCATLT*				

ORF24ng and ORF24-1 show 96.1% identity in 307 aa overlap:

50	orf24-1.pep	10	20	30	40	50	60
	orf24ng	MRTAVVLLILIMPAASSAMPEMVCAGVSPGTAIISKPTTEQTAVMASSLSVSTPASAAA					
55	orf24-1.pep	70	80	90	100	110	120
	orf24ng	IIPSSSETGINAPLKPPTALEAIMPPFFTASFNSAKAAVVPVQTLKPISSRMRAATESP					
60	orf24ng	70	80	90	100	110	120
		IIPSSSETGINAPLKPPTALEAIMPPFFTASFNSAKAAVVPVQTLKPISSRMRAATESP					

-390-

		130	140	150	160	170	180
	orf24-1.pep	TAGVGASDKSRIPNGIFSI FEASRPMSSPTRVILKAVFTTSATSVNVVASEFSNAAFTT					
	orf24ng	TAGVGASDKSRMPNGIFSI FEASRPMSSPTRVILKAVFTTSATSVRLTASEFSNAAFTT					
5		130	140	150	160	170	180
	orf24-1.pep	PGPDTPTLITASASPEPKNAPAINGLSSTALQNTTILAQPKPSGVISAVRLTVSPASLTA					
	orf24ng	PGPDTPTLITASASPEPWNAPAINGLSSTALQNTTILAQPKPSGVISAVRLMVSPASLTA					
10		190	200	210	220	230	240
	orf24-1.pep	SILIPARVLPILMELHTISVVFIA SGMERINTSSGDIPTCTNAEKPPFKDTPMALAALS					
	orf24ng	SILIPARVLPILMELHTISVVFIA SGTERINTSSGDIPTCTSAEKPPFKDTPMALAALS					
15		250	260	270	280	290	300
	orf24-1.pep	SILIPARVLPILMELHTISVVFIA SGMERINTSSGDIPTCTNAEKPPFKDTPMALAALS					
	orf24ng	SILIPARVLPILMELHTISVVFIA SGTERINTSSGDIPTCTSAEKPPFKDTPMALAALS					
20		250	260	270	280	290	300
	orf24-1.pep	KVCATLTx					
	orf24ng	KVCATLTx					

Based on this analysis, including the presence of a putative leader sequence (first 18 aa – double-underlined) and putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 81

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 681>:

30	1	..ACCGACGTGC	AAAAAGATT	GGTCGGCGAA	CAACGCAAGT	GGGCGCAGGA
	51	AAAAATCAGC	AACTGCGGAC	AAGCGCGCGC	GCAGGCGAGAC	CGGCGGAAAT
	101	ACGCGGAATA	CCTCAAGGTG	CAATGCGACA	CGCGSATGAC	GCGCGAACGG
	151	ATACAGTATC	TTGCGCGCTA	TTCCATCGAT	TAG	

This corresponds to the amino acid sequence <SEQ ID 682; ORF25>:

35	1	..TDVQKELVGE	QRKWAQEKIS	NCRQAAAQAD	RQEYAEYLKL	QCDTRMTRE
	51	IQLRGSYID	*			

Further work revealed the complete nucleotide sequence <SEQ ID 683>:

	1	ATGTATCGGA	AACTCATGTC	GCTGCGTTT	GCCTGCTGC	TTGCCGCTTG
	51	CGGCAGGGAA	GAACCGCCCA	AGGCATTGGA	ATGCGCCAAC	CCCGCGTGT
40	101	TGCAAGGCAT	ACGCGGCAT	ATTCAAGAAA	CGCTCACGCA	GGAAAGCGGT
	151	TCTTTTCGCG	GCGAAGACGG	CAGGCAGTTT	GTCATGCGC	ACAAAATTAT
	201	CGCGCGCGC	TACGTTTTCG	CGTTTCTTTT	GGAACACGCT	TGGAAACGCT
	251	AGGAAGCGCG	GCGCAGTTTC	TGTATCGCG	ATTGAAACAT	TACCGTGGCG
	301	TCTGAAACGC	TTGCGATGTC	CAAGGCAAA	AGCCCGCTGT	TGTACGGGGA
45	351	AACTCGTTTG	TGCGATATTG	TGCGGAGAAA	GACGGCGGCG	AATGTGAGT
	401	TTAAGACGG	CGTATTTCAG	CGAGCGCTCC	GCTCTCTGCG	CGTCAAAAG
	451	GCTCAAGCG	CATTTCGGA	CACACGGCT	GATATGCGCG	CGCAAAACCT
	501	CTCTGCGCG	CTGCTGCTTT	ACGCGGTGAA	GAGCATCGTG	ATGATAGACG
	551	GCAAGGCGGT	GAAGAAAGAA	GACGCGGTCA	GGATTTTGAG	CGGAAAGGCC
50	601	CGTGAAGAG	AACCGTCCAA	ACCCACGCC	GAGACATTTT	TGGAACACAA
	651	TGCCGCGCG	GGCGTTCGCG	GCGTACCCCA	AGCGCGAGAA	GGCGGCGCG
	701	AACCGGAAAT	CCTGCATCCT	GACGACGGCG	AGCGTGCCGA	TACCGTTACC
	751	GTATCAAGGG	GCGAAGTGG	AGAGCGCGCG	GTACAAAACC	AGCGTGCGGA
	801	ATCCGAAATT	ACCAAACTTT	GGGAGAGACT	CGATACCGAC	GTGCAAAAG
55	851	AGTTGGTCGG	CGAACAACGC	AGTGGCGCG	AGGAAAAAT	CAGCAACTGC
	901	CGACAAGCG	CGCGCGAGCG	AGACCGGCG	GAATACGCG	AATACCTCAA
	951	GCTGCARTGC	GACACGCGGA	TGACGCGCGA	ACGGATACAG	TATCTTGGCG
	1001	GCTATTCCAT	CGATTAG			

This corresponds to the amino acid sequence <SEQ ID 684; ORF25-1>:

```

1 MYRKLIALPF ALLLAACGRE EPPKALECAN PAVLQIGIRN IQETLTQEAR
51 SFAREDDGRQF VDADKIIAAA YGLAFSLEHA SETQEGGRFT CIADLNITVP
101 SETLADAKAN SPILYGETAL SDIVRQKTGG NVEFKDGVLT AAVRFLPVKD
151 GQTAFDVNTV GMAAQTLSAA LLPYGVKSIV MIDGKAVKE DAVRILSGKA
201 REEPEPSKPTP EDILEHNAAG GDAGVPQAAE GAPEPEILHP DDGERADTVT
251 VSRGEVEEAR VQNQRAESEI TKLWGLDLD VQKELVGQR KWAQEKISNC
301 RQAAQADQRQ EYAAYLKLQC DTRMTREIRI QYLRGYSID*

```

Computer analysis of this amino acid sequence gave the following results:

10 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF25 shows 98.3% identity over a 60aa overlap with an ORF (ORF25a) from strain A of *N.*

meningitidis:

```

15 orf25.pep          TDVQKELVGEQRKWAQEKISNCRQAAQAD
orf25a          250 260 270 280 290 300
VTVSRGEVEEARVQNQRAESEITKLWGLDLDVQKELVGEXRRKWAQEKISNCRQAAQAD

20 orf25.pep          40 50 60
RQEYAYLYKLQCDTRMTREIRIQYLRGYSIDX
|||||
orf25a          310 320 330
RQEYAYLYKLQCDTRMTREIRIQYLRGYSIDX

```

The complete length ORF25a nucleotide sequence <SEQ ID 685> is:

```

25 1 ATGTATCGGA AACTCATTGC GCTGCGGTT GCCCTGCTGC TTGCGCGTTG
51 CGGCAGGGAA GAACCGGCCA AGGCATTGGA ATGCGCCAC CCGCGCTGT
101 TGCAANGCAT ACGCNGCAAT ATTCAGGAAA CGCTCAACGA GGAAGCGCGT
151 TCTTTCGCGC GCGAAGACNG CANGCAGTTT GTCGATGCGC ACNAAATTAT
201 CGCCGCCGCC TANGNTNNGN NGMNTICTTT GGAACACGGT TCGGAAACGC
301 AGGAAGCGCG GCGCAGCTTC TGNTGCGCG ATTTGAACAT TACCGTCCCG
351 TCTGAAACGC TTGCGCATGC CAGGCGAACC AGCCCTCTGC TGTACGGGGA
401 TTAAGACGCG CGTATTGACG GCGACCGCTC GCTTCTACC CGTCAAAGAC
451 GGTACAGANG CATTTGTCGA CAACACGGTC GGTATCGGG CGCAACCGT
501 GTCTGCGCGC TTGCTGCTTT ACGCGGTGAA GAGCATGGT ATGATAGACG
551 GCAAGCGCGT AAAAAAGAAA GACGCGGTCA GGATTNTGAG CNGANAAGCG
601 CGTGAANAAG AACCTGTCGA ANCCNNGCCC GAAGACATTT TGGACATAAA
651 TGCCGCGCGA GGGGATGCAG ACGTACCCCA AGCCGGAGAA GACGCGCCCG
701 AACCGGAATV CCTGCATCCT GACGACGGCG AGCGTGCGCA TACCGTTACC
40 751 GTATCACGGG GCGAATGGA AGAGGCGCGN GTACAAAACC AGCGTGCGGA
801 ATCCGAATTT ACCAAACTTT GGGGAGGACT CGATACCGAC GTGCAAAAAG
851 AGTTGTGTCG CGAANAACGC AAGTGGGCGC AGGAAAAAAT CAGCAACTGC
901 CGACAAGCGC CCGCGCAGGC AGACCGCGAG GAATACGCGG AATACCTCAA
951 GGTGCAATGC GACACGCGGA TGACGCGGGA ACGGATACAG TATCTTCGCG
45 1001 GCTATTCCAT GCAATTAG

```

This encodes a protein having amino acid sequence <SEQ ID 686>:

```

50 1 MYRKLIALPF ALLLAACGRE EPPKALECAN PAVLQIXIRN IQETLTQEAR
51 SFAREDDXXQF VDADKIIAAA XXXXXSLEHA SETQEGGRFT CXADLNITVP
101 SETLADAKAN SPILYGETAL SDIVRQKTGG NVEFKDGVLT AAVRFLPVKD
151 GQXAFDVNTV GMAAQTLSAA LLPYGVKSIV MIDGKAVKE DAVRILSXXA
201 REXEPEPSKXFP EDILEHNAAG GDADVPQAGE DAPEPEILHP DDGERADTVT
251 VSRGEVEEAR VQNQRAESEI TKLWGLDLD VQKELVGEXR KWAQEKISNC
301 RQAAQADQRQ EYAAYLKLQC DTRMTREIRI QYLRGYSID*

```

ORF25a and ORF25-1 show 93.5% identity in 338 aa overlap:

```

55 orf25a.pep          10 20 30 40 50 60
MYRKLIALPFALLLAACGRE EPPKALECANPAVLQIXIRNIQETLTQEARSFAREDDXXQF
|||||
orf25-1          MYRKLIALPFALLLAACGRE EPPKALECANPAVLQIGIRNIQETLTQEARSFAREDDGRQF

```

		10	20	30	40	50	60	
5	orf25a pep	VDADXI	7AA	XXXXXXSL	HEASETQEGGR	TFCKAD	LNITVPSETLADAKANSPLLYGETAL	
	orf25-1	VDADKI	7AA	YGLAFSL	HEASETQEGGR	TFCKAD	LNITVPSETLADAKANSPLLYGETAL	
			70	80	90	100	110	120
10	orf25a pep		130	140	150	160	170	180
	orf25-1		130	140	150	160	170	180
			130	140	150	160	170	180
15	orf25a pep		190	200	210	220	230	240
	orf25-1		190	200	210	220	230	240
			190	200	210	220	230	240
20	orf25a pep		250	260	270	280	290	300
	orf25-1		250	260	270	280	290	300
			250	260	270	280	290	300
25	orf25a pep		310	320	330	339		
	orf25-1		310	320	330	339		
			310	320	330	339		

ORF25 shows 100% identity over a 60aa overlap with a predicted ORF (ORF25ng) from

35 *N.gonorrhoeae*:
 orf25.pep T D V Q K E L V G E Q R K W A Q E K I S N C R Q A A A Q A D 308
 |||||
 orf25ng V T V S R G E V E E A R V Q N Q R A E S E I T K L W G L D T D V Q K E L V G E Q R K W A Q E K I S N C R Q A A A Q A D 308
 |||||

40
 orf25.pep R Q E Y A E Y L K Q C T R M T R E R I Q Y L R G Y S I D 60
 |||||
 orf25ng R Q E Y A E Y L K Q C T R M T R E R I Q Y L R G Y S I D 338

The complete length ORF25ng nucleotide sequence <SEQ ID 687> is:

45	1	ATGATATCGGA	AACATCATTCG	GGTGGCGGTTT	GGCGTCGCTG	TTGAGAGCTG
	51	CGCGAGGGGA	AGCGATGCCCA	AGGCTATTGGA	ATGTCGCACAC	CCGCGCTGTG
	101	TGCGAGCATAT	AGCGCGGCAT	AGTCAAGAGCA	GCTTCACGAC	GGAGCGCGGT
	151	TCCTTTCCGCG	CGCAGACACGG	CAGGACSTMTT	GGTATGTCGG	ACAAATTTAT
50	201	CGGACGCGCGC	TGCAGTTTGG	CGTTTCTTCTT	GGACACACGT	TGCGAAACGC
	251	AGCGAGCGGCG	GCGACAGCTG	TGTATTGCGG	TATTTAGCAAT	TACCGTGGCG
	301	TCTGAAACCG	TGCGCGATGC	CGAGGCAAAAC	AGGCCCTCTG	TGTATTGGGA
	351	AACGTCTTTTG	CGACACATCG	TGCACGAGCA	GACGGGCGCG	AATGTGCTGG
55	401	TTAAAGACGC	CGTATTGACG	CGACCCCTGC	CGTCTTCCGC	GCCCAAAAGC
	451	CTCGGACGG	CATTATTACG	CACACGGCT	GGTATGGCGA	CGCAACCGCT
	501	GTCGCGCGCG	TGTCTGCTCT	ACGGCGGTGA	AGCATCTGTG	ATGATGACGC
	551	GCATGGCGGT	GACAAAGAA	GCGCGGTCA	GGGTTTGG	CGCGAAACCG
60	601	TGCGAAGAGC	AACCGTCCAA	ACCACACCCC	GAAGACATCT	TGGACACACG
	651	CGCGCGCGCG	GCGGATGCGG	GGGTACCCCA	GCGCCAGAA	GGCGCACCGG
	701	AACCGCAAGT	CTCGATCTCG	GACGAGCGTG	AGCGTCAGCA	TACCGTTTGA
	751	GTATCATGGG	CGCATATGCC	AGAGGCGCGC	GTACAAACGC	ACGCTGCGGA
65	801	NTCGGAATTC	CGCAAACTAT	GGGAGGAGAT	CGATTCGCGG	GTGCAAAAGC
	851	AGTTGGTGGG	CGGCGGAGCG	AGTCGCGGCG	CGGCGGCGCG	CGGCGGCGCG
	901	CGCAAGCGTC	CGCGCACATG	AGCGAGCGAC	GAATTCGCGG	AATTAACCTAA
	951	CGTCTCATGC	GACACGCGGA	TGACGCGCGA	ACggatACAG	TATCTTCGCG
1001	GCATTTCTAT	CGATTGAG				

65 This encodes a protein having amino acid sequence <SEQ ID 688>:

-393-

		1	MYRKLIALPF	ALLLAACGRE	EPPKALECAN	PAVLQDIRGS	IQETLTQEAR		
		51	SFAREDGRQF	VDADKIIAAA	YGLAFSLEHA	SETQEGGRTF	CIADLNITVP		
		101	SETLADAEAN	SPLLYGETSL	ADIVQKGTGG	NVEFKDGVLT	AAVRFLPAKD		
5		151	ARTAFIDNTV	GMATQTLTSA	LLPYGVKSIV	MIDGKAVKTR	DAVRVLSGKA		
		201	REEEPSKPTP	EDILEHNAAG	GDAGVPQAAE	GAPEPEILHP	DVERADTVT		
		251	VSRGEVEEAR	VQNGRAPSEI	TKLWGGLDTP	VQKELVGEQR	KWAQEKISNC		
		301	RQAAAQADRDQ	EYAEYKLQCC	DTMTRERITQ	YLRGYSID*			
ORF25ng and ORF25-1 show 95.9% identity in 338 aa overlap:									
10	orf25-1.pep	10	20	30	40	50	60		
	orf25ng	10	20	30	40	50	60		
15	orf25-1.pep	70	80	90	100	110	120		
	orf25ng	70	80	90	100	110	120		
20	orf25-1.pep	130	140	150	160	170	180		
	orf25ng	130	140	150	160	170	180		
25	orf25-1.pep	190	200	210	220	230	240		
	orf25ng	190	200	210	220	230	240		
30	orf25-1.pep	250	260	270	280	290	300		
	orf25ng	250	260	270	280	290	300		
35	orf25-1.pep	310	320	330	339				
	orf25ng	310	320	330					

- 45 Based on this analysis, including the presence of a predicted prokaryotic membrane lipoprotein lipid attachment site (underlined) in the gonococcal protein, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

- ORF25-1 (37kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 16A shows the results of affinity purification of the GST-fusion protein, and Figure 16B shows the results of expression of the His-fusion in *E.coli*. Purified His-fusion protein was used to immunise mice, whose sera were used for Western blot (Figure 16C), ELISA (positive result), and FACS analysis (Figure 16D). These experiments confirm that ORF25-1 is a surface-exposed protein, and that it is a useful immunogen.

Figure 16E shows plots of hydrophilicity, antigenic index, and AMPHI regions for ORF25-1.

Example 82

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 689>

```

5      1  ATGCAGCTGA  TCGACTATTC  ACATTCAATT  TTCTCGGTTG  TGCCACCCCTT
      51  TTTGGCACTG  GCACCTGGCG  TCATTACCGG  CGCGTACTG  CTGCTTTAG
      101 GCATCGGTAT  TCTGGTCGGC  GTTGCCTTTT  TGGTCCGGGG  CAAACCCGTC
      151 GACGGCTCTGA  CACACCTGAA  AGACATGGTC  GTCGGCTTGG  CTTGGTCAGA
      201 CGSgATTGG  TCGCTGGGCA  AACCAAAAAT  CTTGGTTTTC  CTGATACTTT
      251 TGGGTATTTT  TACTTCCCTG  CTGACCTACT  CGGGCAGCAA  T.....

10     851  .....AC  TTGCTGGTA
      901  TTGCGCGGCA  CTTGGGCGGT  CTTTGGCCTC  GTTCTCTGCA  CGCTCGGCAC
      951  GATTAAAAACC  GCCGACTATC  CCAAAGCGGT  TTGCGAGGGT  GCGAAATCTA
      1001 TGTTCCGGCG  AATCGCCATT  TTAATCTCTG  CTTGGCTCAT  CAGTACGGTT
      1051 GTCGGCGAAA  TGCACACGGG  CGATTACCTC  TCCACACTGG  TTGGGGCCAA
      1101 CATCCATCCC  GGGTTCCTGC  CGGTCACTCT  TCTCTGCTC  GCGAGCGTGA
      1151 TGGCGTTTTC  CACAGGCACA  AGCTGGGGGA  CGTTCGGCAT  TATGCTGGCG
      1201 ATTGCCCGCG  CCATGGCGGT  CAAAGTCGAA  CCGCGCTGTA  TTATCCCGTG
      1251 TATGTCGCGA  GTAATGGCGG  GGGCGGTATG  CGGCGACAC  TGCTGCCCA
      1301 TTTCCGACAC  GACCATCTCG  TCGTCCACGG  GCGGCGCTG  CAACACATC
      1351 GACCACGTTA  CCTCGCAACT  GCGTTACCGT  TTAACCGTTG  CCGCCCGCG
      1401 CGCATCGGGC  TACCTCGCAT  TGGTCTGAC  AAAATCCGGG  CTGTGGGCT
      1451 TTGGCAGCAC  AGGCATTGTA  TTGGCGGTGC  TGATTTTCT  GTTGAAAGAT
      1501 AAAAAA..

```

25 This corresponds to the amino acid sequence <SEQ ID 690; ORF26>:

```

      1  MQLIDYSHSF  FSVVPPFLAL  ALAVITRRVL  LSLGIGLXX  VAFLVGGNPV
      51  DGLTHLKDMV  VGLAWSDXDW  SLGPKILVF  XILLGIFSL  LTYSGSN...

30     251  .....TSLV
      301  FGGTCGVFAV  VILCTLGIKT  ADYKRAVWG  AKSMFGAIAL  LILAWLISTV
      351  VSEMHGDTL  STLVAQHHP  GLFVILFLL  ASVNAATGT  SWGTFLHLP
      401  TAAAMVRYE  PALIIPCMA  VMAGAVGDH  CSPISDITLL  SSTGARGNHI
      451  DHVTSQLPYA  LTVAAAAAG  YLALGLTKSA  LLGPTTGTIV  LAVLIFLLKD
      501  KK..

```

35 Further work revealed the complete nucleotide sequence <SEQ ID 691>:

```

      1  ATGCAGCTGA  TCGACTATTC  ACATTCAATT  TTCTCGGTTG  TGCCACCCCTT
      51  TTTGGCACTG  GCACCTGGCG  TCATTACCGG  CGCGTACTG  CTGCTTTAG
      101 GCATCGGTAT  TCTGGTCGGC  GTTGCCTTTT  TGGTCCGGGG  CAAACCCGTC
      151 GACGGCTCTGA  CACACCTGAA  AGACATGGTC  GTCGGCTTGG  CTTGGTCAGA
      201 CGCGGATTGG  TCGCTGGGCA  AACCAAAAAT  CTTGGTTTTC  CTGATACTTT
      251 TGGGTATTTT  TACTTCCCTG  CTGACCTACT  CGGGCAGCAA  TCAGGCGTTT
      301 GCGGACTGGG  CAAACCGGCA  CATTAAAAAC  CGGCGCGGCG  CGAAAAATGCT
      351 GACGCGCTGC  CTCGTGTTGG  TAACCTTTAT  CGAGCACTAT  TTCCACAGTC
      401 TCGCGCTGGG  TCGGATTGCC  CGCCCGGTGA  CGGCAAGTGT  TAAAGTTTCG
      451 CGGACCAAC  TGGCTACAT  CCTGACTCC  ACTGCGGCT  CTATGTCG
      501 GCTGATGCC  GTTTCAAAGT  GGGGCGGCTC  GATTATCGCC  AGCCTTGGCG
      551 GACTGCTCGT  TACCTACAAA  ATCACCGGAT  ACACGCGAT  GGGGAGTTT
      601 GTCGCCATGA  GCGTGATGAA  CTAATAGGCA  CTGTTTGCCC  TGATTATGTT
      651 GTTCGTGCTG  GCATGGTTTT  CCTTGACAT  CGGCTCGGAT  GCAAGTTTCG
      701 AACCAAGCGC  GTTGAACGAA  GCCACGATG  AAACCTCGGT  TTACAGCGCT
      751 ACCAAAGGTC  GTGTTTAGCG  ACTGATTAAT  CCGCTTTTGG  CCTTAATCGC
      801 CTCACCGGTT  TCGGCCATGA  TCTACACCGG  CGCGCAGGCA  AGCGAAACCT
      851 TCAGCATTTT  GGGGGCATTT  GAAACACGCG  ACCTAATAC  TTGCTGGTGA
      901 TTGCGCGGCA  CTTGGGCGGT  CTTGCGGCT  GTTCTCTGCA  CGCTCGGCAC
      951 GATTAAAAACC  CGCGACTATC  CCAAAGCGGT  TTGGCAGGGT  GCGAAATCTA
      1001 TGTTCCGGCG  AATCGCCATT  TTAATCTCTG  CTTGGCTCAT  CAGTACGGTT
      1051 GTCGGCGAAA  TGCACACGGG  CGATTACCTC  TCCACACTGG  TTGGGGCCAA
      1101 CATCCATCCC  GGGTTCCTGC  CGGTCACTCT  TCTCTGCTC  GCGAGCGTGA
      1151 TGGCGTTTTC  CACAGGCACA  AGCTGGGGGA  CGTTCGGCAT  TATGCTGGCG
      1201 ATTGCCCGCG  CCATGGCGGT  CAAAGTCGAA  CCGCGCTGTA  TTATCCCGTG
      1251 TATGTCGCGA  GTAATGGCGG  GGGCGGTATG  CGGCGACAC  TGCTGCCCA
      1301 TTTCCGACAC  GACCATCTCG  TCGTCCACGG  GCGGCGCTG  CAACACATC

```



```

1351 GACCACGTTA CCTCGCACT GCCTTACGCC TTAACCGTGG CCGCGCGCGC
1401 CGCATCGGGC TACCTCGCAT TGGGTCTGAC AAAATCCGCG CTGTTGGGCT
1451 TTGGCACGAC AGGCATTGTA TTGGCGGTGC TGATTTTTCT GTTGAAGAT
1501 AAAAAACGCG CCAACGCGCT A

```

- 5 This corresponds to the amino acid sequence <SEQ ID 692; ORF26-1>:

```

1 MQLIDYSHSF FSUVPPFLAL ALAVITRVL LSLGIGILV VAFVVGNGPV
51 DGLTHLKMV VGLAWSGDW SLGPKILVF LLLGIFTSL LTYSGSNQAF
101 ADWAKRHIM RRGAXMLTAC LVFTTIDDY PSLAVGATA RPVTOFKVYS
151 RTKLAVILDS TAAPMCLVMP VSSWGASIA TLAGLLVYK IETVPMGTG
201 VAMSLMNYA LFLIMVFLV AWFSEIDGSM ARFEQALNE AHDETAVSDA
251 TKGRVYALII FVLALIASTV SAMIYTGAAQ SETFSILGAF ENTVDNTSLV
301 FGGTCGVLA VCLTGLTIKT ADYPAKAWOG AKSMFGAIAI LILAWLISTV
351 VGMHTGDYLV STLVAAGNIH GLFPLVILFL ASVMAFATGT SWGTFGIMLP
401 IAAAMAVKVE PALIIPCMSA VMAGAVCGDH CSPISDITIL SSTGARNCHI
451 DHVTSQLPYA LTVAAAAAGS YLALGLTRSA LLFGTGTGIV LAVLIFLFLK
501 KGRANA*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with the hypothetical transmembrane protein HI1586 of *H. influenzae* (accession number P44263)

ORF26 and HI1586 show 53% and 49% amino acid identity in 97 and 221 aa overlap at the

- 20 N-terminus and C-terminus, respectively:

```

Orf26 1 MQLIDYSHSF FSUVPPFLALALAVITRRVXXXXXXXXXXVAVLVGNGPVDGLTHLKMV 60
M+LID+S S+S+VP LA+ LA+ TRRV L +L V
HI1586 14 MELIDFSSSVSIVPALLAILAIATRRVLVSLSAGIIIGSLMSDWIGSAFNYLVKNV 73

```

- 25 Orf26 61 VGLAWSDXDWSLKGPKILVFIXILLGIFTSLTYSGSN 97
V L +D + + I +F +LLG+ T+LLT SGN
HI1586 74 VSLVYADGEIN-SNMNIVFLFLLGLVLTALTYSGSN 109

//

```

Orf26 86 IFTSLTYSGS--NTSLVFGGTCGVFAVVLCTL--GTIKTADYPAKAVQWQASMFQXXXX 141
+F+ L T+ + TSLV GG C + L + + +Y ++ G KSM G
HI1586 299 VFSVLGTFFENTVVGTSLVVGGFCIIITSLLLILDROVSUPEYVRSWIVGIKSMGSAIAI 358

```

- 35 Orf26 142 XXXXXXSTVVMGEMHTGDYLVSTLVAGNIHGLFPLVILFLLASVMAFATGTSGWTFGIMLP 201
+ +VG+M TG YLS+LV+GNI FLPLVIL+L + MAF+TGTSGWTFGIMLP
HI1586 359 LFFAWTINKIVGDMQTGKYLSSLVSGNIPMQFLPVLVFLVLAAMAFSTGTSGWTFGIMLP 418

- 40 Orf26 202 IAAAMAVKVEPALIIPCMSAVMAGAVCGDHCSPISTDTILSSTGARNCHIDHVTSSQXXX 261
IAAAMA P L++PC+SAVMAGAVCGDHCSPISTDTILSSTGA+CNIHDHVT+O
HI1586 419 IAAAMAANAPELLLPCLSAVMAGAVCGDHCSPISTDTILSSTGARNCHIDHVTQLPYA 478

- 45 Orf26 262 XXXXXXXXXXXXXXXXXXXXKALLGFTGTGIVLAVLIFLKKM 302
S L GF T + L V+IF +K +
HI1586 479 ATVATATISIGYIVVGFTYSGLAGFAATVSLVLIIFAVKKR 519

Homology with a predicted ORF from *N. meningitidis* (strain A)

ORF26 shows 58.2% identity over a 502aa overlap with an ORF (ORF26a) from strain A of *N.*

meningitidis:

```

50 orf26.pcp 10 20 30 40 50 60
MQLIDYSHSF FSUVPPFLALALAVITRRVLLSLGIGILXXVAVLVGNGPVDGLTHLKMV
orf26a 10 20 30 40 50 60
MQLIDYSHSF FSUVPPFLALALAVITRRVLLSLGIGILVGVAVLVGNGPVDGLTHLKMV

55 orf26.pcp 70 80 90 99
VGLAWSDXDWSLKGPKILVFIXILLGIFTSLTYSGSNXX-----
orf26a 70 80 90 100 110 120
VGLAWSGDWSLKGPKILVFLLGLIFTSLTYSGSNQAFADWAKRHIKNRGAARKMLTAC

```

-396-

5	orf26.pep	-----
	orf26a	<u>LVFVTFIDDYFHSIAVGAXARPVTDKFKVSRKAIAYILDSTAAPMCVLMVSSSWGASIIA</u>
		130 140 150 160 170 180
10	orf26.pep	-----
	orf26a	<u>TLAGLLVTKITEYTFMGTFVAMSLMNYALFALIMFVVAWFSFDIGSMARFEQAALNE</u>
		190 200 210 220 230 240
15	orf26.pep	-----
	orf26a	<u>AHDETAVSDGSGWRVYALIIIFVLALIASTVSAMITGAQASETFSILGAFENTDVNTSLV</u>
		250 260 270 280 290 300
20	orf26.pep	-----
	orf26a	<u>FGGTCGVFAVVLCTLGTEKTADYPKAVWQGAISMFGAIAIILAWLISTVVVGMHTGDYI</u>
		120 130 140 150 160 170
25	orf26.pep	-----
	orf26a	<u>FGGTCGVIAVVLCTLGTEKTADYPKAVWQGAISMFGAIAIILAWLISTVVVGMHTGDYI</u>
		310 320 330 340 350 360
30	orf26.pep	-----
	orf26a	<u>STLVAGNIHPGLPVILFLASVMAFATGTSWGTFGIMLPITAAAMAVKVEPALIIPCMSA</u>
		180 190 200 210 220 230
35	orf26.pep	-----
	orf26a	<u>STLVAGNIHPGLXVILFLASVMAFATGTSWGTFGIMLPITAAAMAVKVDFPSIIPCMSA</u>
		370 380 390 400 410 420
40	orf26.pep	-----
	orf26a	<u>VMAGAVCGDHCSPISDITTLSSSTGARCNIHDVTSQLPYALTVAASAGSYIALGLTKSA</u>
		240 250 260 270 280 290
45	orf26.pep	-----
	orf26a	<u>VMAGAVCGDHCSPISDITTLSSSTGARCNIHDVTSQLPYALTVAASAGSYIALGLTKSA</u>
		430 440 450 460 470 480
50	orf26.pep	-----
	orf26a	<u>LLGFGTTGIVLAVLIFLLKDKK</u>
		300 310
55	orf26.pep	-----
	orf26a	<u>LLGFGXTGIVLAVLIFLLKDKKRANAX</u>
		490 500

The complete length ORF26a nucleotide sequence <SEQ ID 693> is:

45	1	ATGCAGCTGA	TCGACTATTC	ACATTCATT	TTCTCGGTG	TGCCACCCCT
	51	TTTGGCACTG	GCACTTGGCG	TCATTACCG	CGGGTACTG	CTGTCTTATG
	101	GCATCGTAT	TCTGGTCGGC	GTTGCCCTTT	TGGTCGCGG	CAACCCGCTC
	151	GACGGCTGA	CACACCTGAA	AGACATGSC	GTGCGCTGG	CTTGTCAGA
	201	CGGCGATTG	TCGCTGGGCA	AACCAAAANT	CTTGTCTTC	CTGATACTTT
	251	TGGGTATTTT	TACTTCCCTG	CTGACCTACT	CGGCGACGAA	TCAGCGSTTT
50	301	GCGGACTGG	CAAAACGGCA	CATTAAAAAC	CGGCGCGGG	CGAAAATGCT
	351	GACCGCTGC	CTCGTGTTCG	TAACCTTTAT	CGACGACTAT	TTCCACAGTC
	401	TGCGCGTGG	TGCGNTTGCC	CGCCCGGTGA	CGGACAAATT	TAAGATTTC
	451	CGCGCCAAAC	TGCGCTACAT	CCTGACTCC	ACTGCGGCGC	CTATGTGCGT
55	501	GCTGATGCC	GTTTCAAGCT	GGGGCGGCTC	GATTATCGCC	ACGCTTGCGC
	551	GACTGCTCGT	TACCTACAAA	ATCACCGAAT	ACAAGCCGAT	GGGGACGTTT
	601	GTGCGCATGA	GCTGATGAA	CTATTACGCA	CTGTTTGCCC	TGATTATGTT
	651	GTTCTCGTTC	GCATGTTCT	CTTCGACAT	CGGCTGATG	GCAGTTTTCG
	701	AACAAGCCGC	GTTGAACGAA	GCCACAGATG	AAACTGCGCT	TTCAACAGGC
	751	AGCTGGGGCA	GGGTTTACGC	ATTGATTATT	CCGCTTTTGG	CTTAATCGC
60	801	CTCAACGGTT	TCCGCCATGA	TCTACACCG	TGCACAGGCA	AGCGAAACCT
	851	TCAGCATTTT	GGGTGCATTT	GAAATACCG	ACGTGAACAC	TTGCTGGTGA
	901	TTCCGGGCGCA	CTTGGCGCGT	GCTTGCGCTC	GTCTCTGCA	CGCTCGGCAC
	951	GATTAAATCT	GCCGATTATC	CAAAAGCCGT	TTGGCAGGGT	GGAAATCCCA
	1001	TGTTTCGGGCG	AATCGCCATT	TTAATCCTTG	CCTGGCTCAT	CAGTACGGTT
65	1051	GTCGGCGGAA	TGCACACAGG	CGACTACCTC	TCCACGCTGG	TTCGGGGCAA
	1101	CATCCATCCC	GGCTTCTCTG	CCGTATCTCT	TTTCTGCTC	GCGACGGTGA
	1151	TGGGTTTTCG	CACAGGCACA	AGCTGGGGGA	CGTTGCGCAT	CAGTGTCCCG
	1201	ATTGCGCGCG	CCATGGCGGT	CAGAAGTCTAT	CCCTCACTGA	TTATCCCGTG
	1251	TATGTCCGCC	GTGATGGCGG	GGGCGGTATG	CGGCGACAC	TGCTCGCCCA
70	1301	TTTCGACAC	GACCATCTG	TGCTTCACCG	GCGCGCGCTG	CAACGCATCT

```

1351 GACCACGTTA CNTCGCAACT GCGTTACGCC TTAAACCGTTG COGCCGCGCG
1401 CGCATCGGGN TACCTCGCAT TGGGTCTGAC AAAATCGCGG CTGTGGTGGT
1451 TTGGCANGAC AGGCATTGTA TTGGCGGTGC TGATTTTCT GTTGAAAGAT
1501 AAAAAACGGC CCAACGCGTG A

```

5 This encodes a protein having amino acid sequence <SEQ ID 694>:

```

1 MQLIDYSHSF FSVVPPFLAL ALAVITRRVL LSLGIGILVG VAFVLVGNPV
51 DGLTHLKMV VGLAWSDDGW SLGKPKXLFV LILLGIFSLT LTYSGSNQAF
101 ADNAKRRIKN RRGAKMLTAC LVFVTFIDY FHSILAVGAXA RPTDKFKVS
151 RAKLAYILDS TAAPMCVIMP VSSWGASTIA TLAGLLVITYK ITEVTPMGTG
201 VAMSLNYYA LFAIMVFPV VAWFSFDGSM ARFEQAALNE AHDETAUSDG
251 SWGRVYALII PVLLALIASTV SAMIYTGQAQ SETFSILGAF ENTDVNTSLV
301 FGGTCGVLAV VLCTLGTIKI ADYPKAVWQG AKSMFGAIAI LILAWLISTV
351 VGMHTGDYL STLVAIGNHP GFLXVILFLL ASVMFAFATG SWTFGIMLP
401 IAAAMAVKVD PSLLIPCMSA VMAGAVCGDH CSPISDTTIL SSTGARNHI
15 DHVTSQPLVA LTVAAAAASG YLALGLTKSA LLGFGXTGIV LAVLIFLLKD
501 KKRANA*

```

ORF26a and ORF26-1 show 97.8% identity in 506 aa overlap:

		10	20	30	40	50	60
20	orf26a.pep	MQLIDYSHSF	FSVVPPFLAL	ALAVITRRVL	LSLGIGILVG	VAFVLVGNPV	DGLTHLKMV
	orf26-1	MQLIDYSHSF	FSVVPPFLAL	ALAVITRRVL	LSLGIGILVG	VAFVLVGNPV	DGLTHLKMV
		10	20	30	40	50	60
		70	80	90	100	110	120
25	orf26a.pep	VGLAWSDDGWS	SLGKPKXLFV	LILLGIFSLT	SLTYSGSNQAF	ADNAKRRIKN	RNRGAKMLTAC
	orf26-1	VGLAWSDDGWS	SLGKPKXLFV	LILLGIFSLT	SLTYSGSNQAF	ADNAKRRIKN	RNRGAKMLTAC
		70	80	90	100	110	120
		130	140	150	160	170	180
30	orf26a.pep	LVFVTFIDY	FHSILAVGAXA	RPTDKFKVS	RAKLAYILDS	TAAPMCVIMP	VSSWGASTIA
	orf26-1	LVFVTFIDY	FHSILAVGAXA	RPTDKFKVS	RAKLAYILDS	TAAPMCVIMP	VSSWGASTIA
		130	140	150	160	170	180
35	orf26a.pep	TLAGLLVITYK	ITEVTPMGT	FVAMSLNYYAL	FALIMVFPV	VAWFSFDG	SMARFEQAALNE
	orf26-1	TLAGLLVITYK	ITEVTPMGT	FVAMSLNYYAL	FALIMVFPV	VAWFSFDG	SMARFEQAALNE
		190	200	210	220	230	240
40	orf26a.pep	AHDETAUSDG	SWGRVYALII	PVLLALIASTV	SAMIYTGQAQ	SETFSILGAF	ENTDVNTSLV
	orf26-1	AHDETAUSDG	SWGRVYALII	PVLLALIASTV	SAMIYTGQAQ	SETFSILGAF	ENTDVNTSLV
		250	260	270	280	290	300
45	orf26a.pep	FGGTCGVLAV	VLCTLGTIKI	ADYPKAVWQG	AKSMFGAIAI	LILAWLISTV	VGMHTGDYL
	orf26-1	FGGTCGVLAV	VLCTLGTIKI	ADYPKAVWQG	AKSMFGAIAI	LILAWLISTV	VGMHTGDYL
		310	320	330	340	350	360
50	orf26a.pep	STLVAGNIHP	GFLXVILFLL	ASVMFAFATG	SWTFGIMLP	PIAAAMAVKVD	PSLLIPCMSA
	orf26-1	STLVAGNIHP	GFLXVILFLL	ASVMFAFATG	SWTFGIMLP	PIAAAMAVKVD	PSLLIPCMSA
		370	380	390	400	410	420
55	orf26a.pep	VMAGAVCGDH	CSPISDTTIL	SSTGARNHI	DHVTSQPLVA	LTVAAAAASG	YLALGLTKSA
	orf26-1	VMAGAVCGDH	CSPISDTTIL	SSTGARNHI	DHVTSQPLVA	LTVAAAAASG	YLALGLTKSA
		430	440	450	460	470	480
60	orf26a.pep	LLGFGXTGIV	LAVLIFLLK	KKRANA			
	orf26-1	LLGFGXTGIV	LAVLIFLLK	KKRANA			
		490	500				
65	orf26a.pep						
	orf26-1						

ORF26 shows 94.8% and 99% identity in 97 and 206 aa overlap at the N-terminus and C-terminus, respectively, with a predicted ORF (ORF26ng) from *N. gonorrhoeae*:

10	orf26.pep	MQLDYSHSFFSVVPPFLALALAVITRRVLLSLGTGLXKVAFLVGGNPVDDLTHLKDMV		60
	orf26.ng	MQLDYSHSFFSVVPPFLALALAVITRRVLLSLGTGLVGVAFVGGNPVDDLTHLKDMV		60
15	orf26.pep	VGLAWSDDXDWSLGKPKILVFXXILGIFTSLITYSGSN		97
	orf26.ng	VGLAWADGDWSLGKPKILVFLILGIFTSLITYSGSNQAFADAKRHKRCGAKMLTAC		120
		//		
20	orf26.pep		TSLVFGGTCGVFAVVLCTLGITKTADYPKA	326
	orf26.ng	ASTVSAMIYTAGAASSETFSILGAFENTDVNTSLVFGGTCGVLAIVLCTFGITKTADYPKA		326
25	orf26.pep	VWQAKSMFGAIAAILLAWLLSTVVGEMHTGDYLSLTVAGNIHPGLEVLFLPLASWAF		386
	orf26.ng	VWQAKSMFGAIAAILLAWLLSTVVGEMHTGDYLSLTVAGNIHPGLEVLFLPLASWAF		386
30	orf26.pep	ATGTSWGTFGIMLPAAAAVMAVKVEPALIIPCMAVMAVAGVGDHCSPISDTTLSSTGAR		446
	orf26.ng	ATGTSWGTFGIMLPAAAAVMAVKVEPALIPCMAVMAVAGVGDHCSPISDTTLSSTGAR		446
30	orf26.pep	CNHIDHVTSQLPYALTVAARAAAGYLAGLGLTKALLFGTGTGIVLAVLFLPKKK		502
	orf26.ng	CNHIDHVTSQLPYALTVAARAAAGYLAGLGLTKALLFGTGTGIVLAVLFLPKKKRADV		506

The complete length ORF26ng nucleotide sequence <SEQ ID 695> is:

35	1	TGCGACGATG	TGACTACTTC	ACATTACTTC	TTCTCGGCTTG	TGCGACCGTT
	51	TTTGGACACTG	GCATGTGCGC	TGATTACCGC	CCGGCGACTG	CTGCTTTTAG
	101	GCATCGGTAT	TTTGGTGGCG	GTGTCCTTTT	TGGTGGCGCG	CACCCCGAGT
	151	CGCGGTATG	CACACCTGCA	AGACATGGTC	GTGCGCTTGG	CTTGGCGCAG
	201	CGCGGTATGG	CTGCTGGGCA	AACCAAAAT	CTTGGTTTTC	CTGATACTTT
40	251	TGGCGATTTT	CACCTCACTG	CTGACCTACT	CGCGASGCA	TACGGGGTCA
	301	GCGCGCTGGG	CAAAAAGGCG	TAATAAAAAC	GTGCGCGCG	CGAAATAGCT
	351	GCGCGCTGCG	CTATGTTTGG	CATTCTTTAT	CGACGCACTG	TTTGCAGGCT
	401	TGCGGCTGGC	CGTGTATGAC	CGCGCGCTTA	TAAAGTGGCT	TAAAGTGGCT
	451	GCGCGCAATG	CGTGCATGAC	CTTGCATGCT	ACTCGCTGCG	CGATAGCGCT
	501	CGTATGCGC	GTTTCAAGCT	GGGCGCGGTC	GATTCGCGC	ACGCTGGTGC
45	551	GATTCGCTGT	TACATACAAA	ATTACCGAT	ACGACCCGAT	GGGCGACGCT
	601	GTGCGCATGA	GCCTGATGAA	CTTATACCG	CTGTTTGGCC	TGATTATGGT
	651	ATTTCGTGCT	CGTGGTTTCT	CTTCCGACAT	CGCGCTGATg	gCGCGTCTG
	701	AACAGCGCTG	GTGGAACGNA	jcccttagagc	aaacgcgcgc	lTCAGACTCg
	751	ACCAAGGTTT	GTGTTTAGCT	ATTGATTATT	CCCGTGTCT	CTTTATTCGC
50	801	TCGCAAGGTC	TGCGGCATGA	TCTACACGGG	GGCGAGGGA	AGCGAAACCT
	851	TGCGATTTT	GGGGGCAATT	GAATATACCG	CGTGGAGGCT	CGTGGAGGCT
	901	TTGCGGCGCT	CGCGGCGCT	CGTGAACGCT	GTGAGAGGCT	GTGAGAGGCT
	951	GTTTAAACCT	CGGCTGTATC	CGTGAACGCT	GTGAGAGGCT	GTGAGAGGCT
55	1001	TGTTGCGCGC	ATGCGCATCT	TTAATCTCTG	CTCGGCTCAT	CAGTACGGTT
	1051	GTGCGCGAAA	TGCAACACGG	GAGCTACTCT	TCCACAGGCT	TGCGGGCGAA
	1101	CATTCATCTC	GGGTTCTCTG	CGCTCATCTC	CTTCTCTGCT	GCCACGGTGA
	1151	TGCGGTTTTC	CACAGGCACA	AGTCTGGGGA	CTTCTGGCAT	TATGCTGCGCG
	1201	ATTGCGCGCG	CATCGCGCGT	CMAAGTCGAA	CCGCGCGCTG	TATatccctga
	1251	TATGTGCGCA	GTAATGGGCG	GGCGGCTATG	GGCGACAC	TGTTGCGCCA
60	1301	TCTCGCAACG	GACATCTCTG	TGCTTCACCG	CGCGCGCGTG	CACACACATG
	1351	GACGCGCTA	CTGCGCAACT	GAGTATGCG	GGGCGGCTG	GGGCGGCGG
	1401	GCATCTGCTG	CGGATGCTG	GGGCGGCTG	AGATGATG	CTTGGGCTG
	1451	TGTCGACGCG	CGGATGTGTA	TGGCGGCTG	CTGATTTTCT	GTGGAAGATG
	1501	AAAAAGCGCT	CGCGGTTTGA	A		

65 This encodes a protein having amino acid sequence <SEO ID 696>:

-399-

	1	MQLIDYSHSF	FSVVPFFLAL	ALAVITRRVL	LSLGIGILVG	VAFVLGGNPV	
	51	DGLTHLKDMV	VGLAWADGDW	SLGKPKILVF	LILLGIFTSL	LTYSGSNOAF	
	101	ADWAKRHIN	ROGAKMLTAC	LVFVTFIDDD	FHSLAVGAIA	RPVTDKFKVS	
5	151	RAKLAYILDS	TASPMCVLMP	VSSWGASIA	TLAGLLVTVK	ITEYTPMGTF	
	201	VAMSLMNYIA	LFALIMVFVV	AWFSFDGSM	ARFEQARLNE	AQDETAASDA	
	251	TGRVYALII	PLALIASTV	SAMIYTGAA	SETFIIIGAF	ENTDVNTSLV	
	301	FGGTGGLAV	VLCTFETIKT	ADYPKAVWQG	AKSMFEGAI	LILAWLISTV	
	351	VGEMHTGDYL	STLVAGNIHP	GFLPVILELL	ASVMAFATGT	SWGTFGIMLP	
	401	IAAAMAVKVE	PALIIPCMSA	VMAGAVCGDH	CSPISDITLL	SSTGARCNIH	
10	451	DHVTSQLPYA	LTVAASASG	YLALGLTKSA	LLGFGTTGIV	LAVLIFLLKO	
	501	KKRADV*					
ORF26ng and ORF26-1 show 98.4% identity in 505 aa overlap:							
		10	20	30	40	50	60
15	orf26-1.pep	MQLIDYSHSF	FSVVPFFLAL	ALAVITRRVL	LSLGIGILVG	VAFVLGGNPV	DGLTHLKDMV
	orf26ng	MQLIDYSHSF	FSVVPFFLAL	ALAVITRRVL	LSLGIGILVG	VAFVLGGNPV	DGLTHLKDMV
		10	20	30	40	50	60
		70	80	90	100	110	120
20	orf26-1.pep	VGLAWSGDGWS	LGKPKILVFL	ILLGIFTSL	LTYSGSNOAF	ADWAKRHIN	KRAGAKMLTAC
	orf26ng	VGLAWADGDWS	LGKPKILVFL	ILLGIFTSL	LTYSGSNOAF	ADWAKRHIN	KRAGAKMLTAC
		70	80	90	100	110	120
25	orf26-1.pep	LVFVTFIDDD	FHSLAVGAIA	RPVTDKFKVS	RKLAIVLDST	AAAPMCVLM	PVSSWGASIA
	orf26ng	LVFVTFIDDD	FHSLAVGAIA	RPVTDKFKVS	RKLAIVLDST	AAAPMCVLM	PVSSWGASIA
		130	140	150	160	170	180
30	orf26-1.pep	TLAGLLVTVK	ITEYTPMGTF	VAMSLMNYIA	LFALIMVFVV	AWFSFDGSM	ARFEQARLNE
	orf26ng	TLAGLLVTVK	ITEYTPMGTF	VAMSLMNYIA	LFALIMVFVV	AWFSFDGSM	ARFEQARLNE
		190	200	210	220	230	240
35	orf26-1.pep	AHDETAASDAT	KGRVYALII	PVLAIASTVS	SAMIYTGAAQ	SETFIIIGAF	ENTDVNTSLV
	orf26ng	AHDETAASDAT	KGRVYALII	PVLAIASTVS	SAMIYTGAAQ	SETFIIIGAF	ENTDVNTSLV
		250	260	270	280	290	300
40	orf26-1.pep	FGGTGGLAVV	VLCTFETIKT	ADYPKAVWQ	GAKSMFEGAI	LILAWLISTV	VGEMHTGDYL
	orf26ng	FGGTGGLAVV	VLCTFETIKT	ADYPKAVWQ	GAKSMFEGAI	LILAWLISTV	VGEMHTGDYL
		310	320	330	340	350	360
45	orf26-1.pep	STLVAGNIHP	GFLPVILELL	ASVMAFATGT	SWGTFGIMLP	IAAAMAVKVE	PALIIPCMSA
	orf26ng	STLVAGNIHP	GFLPVILELL	ASVMAFATGT	SWGTFGIMLP	IAAAMAVKVE	PALIIPCMSA
		370	380	390	400	410	420
50	orf26-1.pep	VMAGAVCGDH	CSPISDITLL	SSTGARCNIH	IDHVTSQ	LPYALITVA	ASASGYLALGLTKSA
	orf26ng	VMAGAVCGDH	CSPISDITLL	SSTGARCNIH	IDHVTSQ	LPYALITVA	ASASGYLALGLTKSA
		430	440	450	460	470	480
55	orf26-1.pep	LLGFGTTGIV	LAVLIFLLKO	KKRADV			
	orf26ng	LLGFGTTGIV	LAVLIFLLKO	KKRADV			
		490	500				
60	orf26-1.pep	LLGFGTTGIV	LAVLIFLLKO	KKRADV			
	orf26ng	LLGFGTTGIV	LAVLIFLLKO	KKRADV			
		490	500				
65	orf26-1.pep	LLGFGTTGIV	LAVLIFLLKO	KKRADV			
	orf26ng	LLGFGTTGIV	LAVLIFLLKO	KKRADV			
		490	500				

In addition, ORF26 ng shows significant homology to a hypothetical *H. influenzae* protein:

-400-

sp|P44263|YF86_HAEIN HYPOTHETICAL PROTEIN HI1586 >gi|1074850|pir||C64037
hypothetical
protein HI1586 - Haemophilus influenzae (strain Rd KW20) >gi|1574427 (U32832) H.
influenzae predicted coding region HI1586 [Haemophilus influenzae] Length = 519
Score = 538 bits (1370), Expect = e-152
Identities = 280/507 (55%), Positives = 346/507 (68%), Gaps = 7/507 (1%)

5 Query: 1 MQLIDYSHSFFSVPPFLAALAVITRRXXXXXXXXXXAFVGGNPVDGLTHLKNV 60
M+LID+S S+S+VP LA+ LA+ TRR L +L V
10 Sbjct: 14 MELIDFSSSVSIVFALLAILAIAIATRVLVLSAGIIGSMLSDWQTGSAFNVLKNV 73

Query: 61 VGLAWADGDWSLGGPKILVFLILLGIFTSLTYSGSNQAFADWAKRIKNCRAKMLTAC 120
V L +ADG+ + I++FL+LLG+ T+LLT SGSN+AFA+WA+ IK R GAK+L A
15 Sbjct: 74 VSLVYADGEIN-SNMNIVLFLLLLGVLTALLTVSGSNRAFAEWAQSRIGRRGAKLLAAS 132

Query: 121 LVFVTFIDDFHSLAVGAIAIRPVTDKFKVSRKALAYILDSTASPMCMVMPVSSWGSASIA 180
LVFVTFIDDFHSLAVGAIAIRPVT+DKFKVSRKALAYILDST+PMCV+MPVSSWGA II
Sbjct: 133 LVFVTFIDDFHSLAVGAIAIRPVTORFKVSRKALAYILDSTAAPCMVMPVSSWGAIIIT 192

20 Query: 181 TLAGLLVITYKITEYTPMGTFVMSIMNYALPALIMVFWVAFVSDIGSMPARFEQAALNE 240
+ GLL TY ITEYTP+G FVMS MN+YA+F+IMVF VA+PSFDI SM R E+ AL
Sbjct: 193 LIGGLLATYSITEYTPIGAFVMSMNFYAI+PSIMVFEVAFVSDIASMVRHEKLAKN 252

25 Query: 241 AQDETAASDATKGRVYALIIIPVLAIASTVMSMIYTGAAQ----SETFSILGAFENTDVN 296
+D+ TKG+V LI+P+L LI +TVS MIYTGAA+ + FS+LG FENT V
Sbjct: 253 TEDQLEETGTGKQVRNLILPILVILIATVSMIYTGAAEAADGKVFSLVGTENTVVG 312

Query: 297 TSLVFGGTCGVL--AVVLCTFGTIKTADYPKAVWGQAKSMFGXXXXXXXXXXSTVVGEM 354
TSLV GG C ++ +++ + +Y ++ G KSM G + +VG+M
30 Sbjct: 313 TSLVVGFGCSIIISTLLIILDRQVSVEYVRSWIGVIGKSMGSAIILFFAWTINKIVGDM 372

Query: 355 HYGDLSTLVAGNIHPGFLPVLIFLLASVMAFATGTSGWTFGIMLPIRAAMAVKVEPALI 414
TG YLS+LV+GNI FLPVILF+L +MAF+TGTSWGTGFMILPIRAAAMA P L+
Sbjct: 373 QTGKYLSSLVSGNIPMQFLPVILFVLGAAMAFSTGTSGWTFGIMLPIRAAAMAAPPELL 432

35 Query: 415 IPCMSAVMAGVCGDHCSPISDITLLSTGACRNHIDHVTSGQXXXXXXXXXXXXXXXXXX 474
+PC+SAVMAGVCGDHCSP+SDITLLSTGA+CNHIDHVT+Q
Sbjct: 433 LPLCSAVMAGVCGDHCSPVSUTLLSTGACRNHIDHVTTLQPYAATATATSIGIYIV 492

40 Query: 475 XXXSALLGFGTTGIVLAVLIFLLKDK 501
S L GF T + L V+IF +K +
Sbjct: 493 GFTYSLAGFAATAVSLVIFIAVKKR 519

Based on this analysis, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*,
45 and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 83

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 697>:

1 ..AAGCAATGGT ATGCGGACGN .AGTATCAAG ACGGAATGG TTATGGTCAA
51 CGATGAGCCT GCCAAATTC TGACTTGGGA TGAAGCGCG CGATTACTCT
50 101 CGGAATCTGTC TATCGGCCAC CATCAACGCA ACGGGTGGT TTTGGAGTGG
151 TATGAAGATG GTTCTAAAAA GAGGCAAGT GTTTATCAGG ATGACAAGTT
201 GGTCAAGAAA ACCCAGTGGG ATAAGGATGG TTATTTAATC GAACCTCTGA

This corresponds to the amino acid sequence <SEQ ID 698; ORF27>:

1 ..KQWYALXSIK TEMVMVNDEP AKILTWDSEG RLLSELSIRH HQRGVNVLEW
55 51 YEDGSKKSEX VYQDKLVKRR TQWKKDGYLI EP*

Further work revealed the complete nucleotide sequence <SEQ ID 699>:

1 ATGAAAAAAT TATCTCGGAT TGTATTTC AACTGCTCTGT TGGGTTTTTC
51 GCGCGCTTTG CCGGCGCAGA CCTATTCTGT TTATTTTAAT CAGAAACGGAA
101 AGCTGACGGC GACGATGTCT TCTGCCGCTT ATATCAGGCA ATATAGTGTG
60 151 GTGGCGGGTA TTGCGCACGC GCAGGATTTT TATTATCCGT CGATGAAGAA

-401-

5
10

201	TATTCTTCGAA	CCCTTATATG	TGTGCTTCAAC	GCACATCAA	TCITTTGTGTC
251	CTACCTCTTCA	AAGTTATGCT	TGTGATTTTAA	TGTCACAGAA	
301	AAATGTCGGG	GGGGCTTTCG	CAAGGTTAAG	CGACCGGGG	AGTGGGTCAA
351	CTAGGTATCC	ACGCGTAAAA	AATCTGCGGT	TCTGGCTTAA	AAAAATGGCT
401	TGAGTGAGGG	TACGGGATAC	CGCTATATCC	GTACGCGGG	CAGGAAGGAA
451	GAATACCGAT	TTAGCAAAA	TAGCAAAAC	GGAATGAGT	AGCAATGGTA
501	TCCAGGATCG	AGTATCAAGA	GGAAATGTT	TGTGCTCAAC	GATGAGCTCG
551	CCAAACATCT	CTGCTGGTAT	CGAAGGSCC	TATTATCTCT	GGAACTGTCT
601	ATCTCCACAC	ATGATACGCA	CGGGTGGTG	TATGAGTGG	ATGAGTGGT
651	TTCTAAAGCA	GGGCTGATG	TTTATAGGA	TACCAATTC	GTCAAGGAAA
701	CCGAGTGGG	TAGGATGCT	TATTTATCG		

This corresponds to the amino acid sequence <SEO ID 700: ORF27-1>:

15

1	MKKLSRIVFS	TVLLGFSAL	PAQYTSVVFN	QNGKLTATMS	SAAYIROQSV
51	VAGIAHQDE	YPPSMKKYSE	PYIVASTQPI	SEVPTLQNGM	LILWHFNGQ
101	KHAGGFSFGK	PDGEWNNYF	KNGLSEGTQV	KNGLSEGTGY	RYRANGGKES
151	ETQFKQNKAN	GVWQWYADG	SKATVMVMN	DEPAKITLW	ESGRLLSELS
201	IRKHQNKAN	GVWYDGSKK	SEAYEDMDK	VKTOWKWDG	YLIE*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

20 ORF27 shows 91.5% identity over a 82aa overlap with an ORF (ORF27a) from strain A of *N. meningitidis*:

		10	20	30		
	orf27.pep	KQWYADSIKTMMVNDPAKITLTDSESG : :				
25	orf27a	LSEGTXRYRRNGGKESEIQFKNQKANGVKQWYAGNIKTEMVMVNDEPAKITLTWDESG 140 150 160 170 180 190				
		40	50	60	70	80
30	orf27.pep	RLLSELIRHHQRNGVVLEWYEDGSKKSEXVQQDKLVKRQTQWDKGYLIPX : :				
	orf27a	RLLSELIRHHXRNGVLWEWYEDGSKKKEAVVQQDKLVKRQTQWDKGYLIPX 200 210 220 230 240				

The complete length ORF27a nucleotide sequence <SEQ ID 701> is:

35	1	GTGCAAAAAAT	TATCTCGGAT	TCGATTTTCA	ACTGTTCCTAT	TGGGTTTTTC
	51	GGCCGCTTTTG	CGCCGCGAGA	NGTATTTTAT	TTATTTTCTAT	CAGAACGGGA
	101	AACTCAGAGCG	GAGGCTGTCT	TCTCGCGCNT	ATATCAGCGCA	ATATAGTGTG
	151	GGGGAGGGTGA	TTCGCGAOCG	CGAGGANTTT	TANTATCTCG	CGATGAAGAA
	201	ATATTTCGAAA	CTCTTATATG	TGCTCTCAAC	GCAATACAAA	TCCTTTGTGC
40	251	TCACCTCGCA	AAACGGTATG	TTGATTTTAT	GGCAATTTAT	NGGTCAAGAA
	301	AAAAATGTCAG	GGGGCTTCAG	CAGGGTAAAG	CAGCGCGGGG	ATGTTGGTCA
	351	CTGGTATTCG	ACCGGTAAAA	TCTGTCCGCT	TATGCCCTAT	AAAAATGGTT
	401	TGAGTGAAGG	TACGGGNTGN	CGCTATACCC	GTACACCGCG	CAGGAAAGGC
	451	TGAATTCAGCT	TTAAATCAGAA	TAGGAGCGGA	TCGGTATGGG	AGCAATAGTG
45	501	TGCGGAGGCG	ATATATCAGA	CGAGAAATGT	TGTGTCGCAT	GATGAGGATG
	551	CGAAATGCGCT	CGGAGGCTGG	CGGCTGCTGG	GGCTGCTGCT	GGCTGCTGCT
	601	ATCTCATCTCT	ATNAGAGCTAA	TGGAGTAGTC	TTAGAGTAGTC	ATAGAGTAGTC
	651	TTCTAAAAAG	ANTAGAGCTG	TTTATCAGGA	TGCTATAGTG	GTCAGGAAAA
	701	CCCATATGGGA	TAAAGATGGT	TATTTTATAT	AGAGTTGCTA	

This encodes a protein having amino acid sequence <SEQ ID 702>:

[illegible]

55 ORF27a and ORF27-1 show 94.7% identity in 245 aa overlap:

10 20 30 40 50 60
orf27a.pep MKKLSRIVFSTVLLGFSAAALPAOXYSYFENONGKLTATXSSAAYIROYSVAEGIAHAXF

10 20 30 40 50 60
orf27-1.pep MKLSRIVFSTVLLGFSAALPAQTYSVYFNONGKLTATMSSAAYIROQYSVVAGIAHAQDF

-403-

[illegible]

Based on this analysis, including the putative leader sequence in the gonococcal protein, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF27-1 (24.5kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 17A shows the results of affinity purification of the GST-fusion protein, and Figure 17B shows the results of expression of the His-fusion in *E.coli*. Purified GST-fusion protein was used to immunise mice, whose sera were used for ELISA, which gave a positive result, confirming that ORF27-1 is a surface-exposed protein and a useful immunogen.

Example 84

The following partial DNA sequence was identified in *N.meningitidis* <SEO ID 705>:

40	1	ATGATGACGTA	CCACGACACC	CGCTGTGGCA	ATGGCGTTC	CGCCATTTTA
	51	TGCGCTGGG	GCGCTGTACG	GCGATCTGCT	CGTATGTCTG	TGGGTTTCG
	101	GTCACACGGG	AACGCACAG	CTGTCGGGTT	TCATATGGCA	CGCGCATGAG
	151	ATGATTGGTG	GTTATGCGCG	AGTGTGCTG	CTGCTGACGC	TGCTGACGCC
	201	CGCTGCCCAT	TGGACACGCG	AGCCGCCACC	CGGGGCGGCG	GTACTCGGTC
45	251	GCGCTGACTA	CTCTTTGGCT	GCGTGGGCGG	ATTGCGCGCT	TTATCCGGGC
	301	TGGGSGTGG	CTGGCAAGCG	GCATACGCG	TACGCTGCT	TTCTGGTAGT
	351	GCGCGGCTTG	CGGCGGCTTG	CGGCGGCTTG	CGACGACG	CGACGACG
	401	TATGTTGCG	TGTTGCGGCT	GTCGCTCTG	CGGCGGCTG	ATGCGGCTTT
	451	CCACGTCGAC	GTGCACACG	GCAACCTAGS	CGGACTCTTG	ACGCGATTGC
	501	AGTGCGGCTT	CTGGATG			

50 This corresponds to the amino acid sequence <SEO ID 706; ORF47>:

1 MKFTKHPVWA MAFRPFYSLA ALYGALSVLL WGFYGTGTHX LSGFYWHAHE
51 MIWGYAGLVV IAFLLTAVAT WTGQPPTRGG VLVGLTIFWL AARIAAFIPG
101 WGAASAGILG TLEFWYGAVC MALPVIRSON ORNYVAVEAL FVLGGTHAAE

151 HVQLHNGNLG GLLSLGQSL VM

Further work revealed the complete nucleotide sequence <SEQ ID 707>:

```

      1 ATGAAATTTA CCAAGCACCC CGTCTGGGCA ATGCGGTTCG GCCCATTTTA
      5 51 TTCGCTGGCG GCTCTGTACG GCGCATTTGC CGTATTGGTG TGGCGTTTCG
      101 GCTACACGGG AACGCACGAG CTGTCGGGTT TCTATTGGCA GCGCATGAG
      151 ATGATTGGGG GTTATGCGGG ACTGGTCGTG ATCGCTTCCG TCGTACACGC
      201 CGTCGCCACT TGGACGGGCG AGCGGCGCAC GCGGGGCGCG GTTCTGGTGG
      251 GCTTGACTAT CTTTTGGCTG GCTGCGCGGA TTGCCGCTT TATCCGGGT
      301 TGGGGTGGT CCGCAGACGG CATACTGGT ACGCTGTTT TCTGGTACGG
      351 CCGGGTGTGC ATGGCTTTGG CCGTTATCCG TTCGAGAA CAACGCAACT
      401 ATGTTGCGGT GTTCGGCGTG TTGCTTTGG GCGGCACGCA TCGGGCGTTC
      451 CACGTCCAGC TGCACACGG CAACCTAGGC GGACTCTGA GCGGATTGCA
      501 GTCGGGCTTG GTGATGGTGT CCGGTTTTAT CGGTCTGATT GGTACGCGGA
      551 TTATTTCGTT TTTTACGTCC AAACGCTTGA ATGTGCCGCA GATTCCCAAT
      601 CCGAAATGGG TGGCGCAGGC TTGCGTGTGG CTGCCCATGC TGACTGCCAT
      651 GCTGATGGCG CACGGTGTGT TTGCTTGGCT GTCTGCCGCT TTGCGCTTTG
      701 CCGCAGGTGT GATTTTTACC GTGCGAGTGT ACGCTGGTGT GTATAAACC
      751 GTGTTGAAG AGCGCATGCT GTGGATTCTG TTTGCCGCTC ATCTGTTTAC
      801 CCGATTGGGG CTGATTGCGG TCGCGCGCTC TTATTTCAAA CCGCGTTTCC
      851 TCAATCTGGG TGTGCATCTG ATCGGGCTCG CGGTATCGG CCGTGTGACT
      901 TTGGGCATGA TGGCGCATAC CGCGCTTGGT CATACGGGCA ATCGGATTGA
      951 TCGCGCGCCC AAAGCCGCTC CCGTTGCGTT TTGCTGTATG ATGCGCGCAA
      1001 CGCGCGTCCG TATGGTTCGC GTATTTTCTT CCGGCACTGC CTACACGCAC
      1051 AGCATCCGCA CCTCTCGGT TTTGTTTGCA CTCGCGCTTT TGGTGTATGC
      1101 GTGGAAGTAT ATTCTCTGGC TGATTCTGCC GCGTTCGGAC GCGAGGCCCG
      1151 GTTGA
  
```

This corresponds to the amino acid sequence <SEQ ID 708; ORF47-1>:

```

      1 MKPTKHPVMA MAFRPFYSLA ALYGALSVLL WFGYTGTHE LSGFYWAHE
      5 51 MIWGYAGLVV IAFLLTAVAT WTGPPTTRGG VLVLGTIWL AARIAAFIPG
      101 WGAASGILG TLFFWYGAVC MALPVRISQN QRNYVAVFAL FVLGGTHAAF
      151 HVQLHNGNLG GLLSLGQSL VMYSGFGLI GTRIIISFTS KRLNVQPIPS
      201 PKWVAQASLV LPMITAMLMA NGVLAWLSAV PAFALGVLET VQVYRWKXFP
      251 VLGEIWLILL FGLVLTSLG LIAVGASTYK PAFNLGVHL IGVGGIVLT
      301 LGMARTALG HTGNPIYPPP KAVPAVPM LM MARTAVRMVA VFSSGTATTH
      351 SIRTSSVLEA LALLVYANKY IFWLIRPRSD GRF*
  
```

Computer analysis of this amino acid sequence predicts a leader peptide and also gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF47 shows 99.4% identity over a 172aa overlap with an ORF (ORF47a) from strain A of *N.*

40 *meningitidis*:

```

      10      20      30      40      50      60
orf47.pep MKPTKHPVMA MAFRPFYSLAALY GALS VLLWFGYTGCTHXLSGFYWHAHEMIWGYAGLVV
      10      20      30      40      50      60
orf47a MKPTKHPVMA MAFRPFYSLAALY GALS VLLWFGYTGCTHXLSGFYWHAHEMIWGYAGLVV

      70      80      90      100      110      120
orf47.pep IAFLLTAVATWTGPPTTRGGVLVGLTI FWLAAARIAAFIPGWGAASGILGTLFFWYGAVC
      70      80      90      100      110      120
orf47a IAFLLTAVATWTGPPTTRGGVLVGLTI FWLAAARIAAFIPGWGAASGILGTLFFWYGAVC

      130      140      150      160      170
orf47.pep MALPVRISQNQRNYVAVFALFVLGGTHAAFIHVQLHNGNLGGLLSGLGSLGLVM
      130      140      150      160      170
orf47a MALPVRISQNQRNYVAVFALFVLGGTHAAFIHVQLHNGNLGGLLSGLGSLGLVMYSGFGLI

      180      190      200      210      220      230      240
orf47a GTRIIISFTSKRLNVQPIPSKWWAQS LVLPMITAMLMAHGVMPWLSAAFAAGVIET
      180      190      200      210      220      230      240
  
```

The complete length ORF47a nucleotide sequence <SEQ ID 709> is:

```

1  ATGAAATTTA CCAAGCACCC CGTTTGGGCA ATGGCGTTCG CCCCGTTTTA
51  TTCACCTGGC GCTCTGTACG GCGGATTGTC CGTATTGCTG TGGGGTTTCG
101 GCTACACGGG AACGCACGAG CTGTCCGGTT TCTATTGGCA CGCGCATGAG
151 ATGATTGGGG GTTATGCGCG ACTGCTCGTC ATCGCCTTCC TGCTGACCCG
201 CGTCGCCACT TGGACGGGCG AGCGGCCGCG CGGGGGCGCG GTTCTGGTCG
251 GCTTGACTAT CTTTGGGCTG GCTCGCGGGA TTGCGGCCTT TATCCCGGGT
301 TGGGGTGGGT CGGCAAGCGG CATACTCGST ACGCTGTTTT TCTGGTACGG
351 CGCGGTGTGC ATGGCTTTGC CCGTATATCG TTGCGAGAAT CAACGCAATT
401 ATGTTTGCCG GTTTCGCGCTG TTGCTCTTGG GCGGTAGCGA CGCGCGCTTC
451 CACGTCACAG TGCACAAACG CAACCTAGGG GGAATCTTGA CGCGATTGCA
501 GTGCGGCTTG GTGATGCTGT CCGGCTTATP CGCTCTGATT GCTACGCGGA
551 TTTATTTGCT TTTTAGCTCC AAACGGTTGA ATCTCCGCGA GATTCCCAAT
601 CGGAATGGGG TGGCGCAGCG TTGCTGTGGG CTCGCCATCG TGACGCCCAT
651 GCTGATGGCG CACGGCGTGA TGCCCTTGGC TTCGCGGGCT TTCGCGTTTG
701 CGGCAGGTGT GATTTTATAC GTGCAGGTGT ACCGCTGGTG GTATAAGCCT
751 GTGTTGAAG AGCCGATGCT GTGGATTCTG TTTGCCGGCT ATCTGTTTAC
801 CGGATTGGGG CTGATTGCGG TCGCGCGGCT TATTTCAAA CCGCGTTTCC
851 TCAATCTGGG TGTGCATCTG ATCGGGGTGCG CGCGTCTGCG CGTGTGACT
901 TTGGGCATGA TGGCGCGTAC CGCGCTCGGT CATACGGGCA ATCCGATTTA
951 TCGCGCGGCC AAAGCCGCTT CCGTTGCGTT TTGGCTGATT ATGGCGGCAA
1001 CGCGCGTCCG TATGTTTGCC GTATTTTCTT CCGGCACGTC CTACACGCAC
1051 AGCATACGCA CCTCTTCGGT TTGTTTGGCA CTGCGCTGTT TGGTGTATGC
1101 GTGGAAGTAT ATTCTTTGGC TGATTCGTCC GCGTTCGGAC GGCAGGCCCG
1151 GTTGA

```

This encodes a protein having amino acid sequence <SEQ ID 710>:

```

1  MKFTKHPVWA MAFRPFYSLA ALYGALSVLL WGFYGTGTHE LSGFYWHAHE
51  MIMGYAGLVV IAFLLTAVAT WTGQPPTRGG VLVGLTIPWL AARIAAFIPG
101  W GASASGILG TLFFWYGAVC MALPVRISQN QRNYVAVFAL FVLGTHAAAF
151  HVQLHNGNLG GLLSGLQSGL VMVSGFIGLI GTRIIISFTFS KRLNVQPIPS
201  PKWVAQASLW LPLMTAMLMA HGVMPWLSAA FAFAPAVIFT VQVYRWYKYP
251  VLKEPMLWIL FAGYLTGILG LIAGVASYFK PAFNLNGVHL IGVGGIGVLT
301  LGHMARTALG HTGNPIYPPP KAVPVAFWLM MAATAVRMVA VFSSGTAYTH
351  SIRTSSVLFA LALLVYAMKY IPWLIRPSD GREG*

```

ORF47a and ORF47-1 show 99.2% identity in 384 aa overlap:

```

10 20 30 40 50 60
orf47a.pep MKFTKHPVWMAFRPFYSLAALYGALSVLLWGFYGTGTHELSGFYWHAHMIMGYAGLVV
101 111 121
orf47-1 MKFTKHPVWMAFRPFYSLAALYGALSVLLWGFYGTGTHELSGFYWHAHMIMGYAGLVV
101 111 121
70 80 90 100 110 120
orf47a.pep IAFLLTAVATWTGQPPTRGGVLVGLTIPWLAARIAAFIPGWGASASGILGTLFFWYGAVC
101 111 121
orf47-1 IAFLLTAVATWTGQPPTRGGVLVGLTIPWLAARIAAFIPGWGASASGILGTLFFWYGAVC
101 111 121
130 140 150 160 170 180
orf47a.pep MALPVRISQNQRNYVAVFALFVLGTHAAAFHVQLHNGNLGGLLSGLQSGLVMSVSGFIGLI
101 111 121
orf47-1 MALPVRISQNQRNYVAVFALFVLGTHAAAFHVQLHNGNLGGLLSGLQSGLVMSVSGFIGLI
101 111 121
190 200 210 220 230 240
orf47a.pep GTRIIISFTFSKRLNVQPIPSPKWVAQASLWLPMLTAMLMAHGVMPWLSAAFAFAGVIFT
101 111 121
orf47-1 GTRIIISFTFSKRLNVQPIPSPKWVAQASLWLPMLTAMLMAHGVMPWLSAAFAFAGVIFT
101 111 121
250 260 270 280 290 300
orf47a.pep VQVYRWYKYPVLKEPMLWILFAGYLTGILG LIAGVASYFKPAFNLNGVHLIGVGGIGVLT
101 111 121
orf47-1 VQVYRWYKYPVLKEPMLWILFAGYLTGILG LIAGVASYFKPAFNLNGVHLIGVGGIGVLT
101 111 121
310 320 330 340 350 360

```

-406-

	orf47a.pep	LGMMARTALGHTGNPIYPPFKAVPVAFWLMMATAVRMVAVFSSGTAYTHSIRTSVSLFA
	orf47-1	LGMMARTALGHTGNPIYPPFKAVPVAFWLMMATAVRMVAVFSSGTAYTHSIRTSVSLFA 310 320 330 340 350 360
5		
	orf47a.pep	LALLVYANKYIPWILIRPRSDGRPGX
	orf47-1	LALLVYANKYIPWILIRPRSDGRPGX 370 380
10		

Homology with a predicted ORF from *N.gonorrhoeae*

ORF47 shows 97.1% identity over 172 aa overlap with a predicted ORF (ORF47ng) from *N.gonorrhoeae*:

15	ORF47	MKPTKHPVKAMAPRPPYSLAALYGLSVLLWGPYGTGTHELSGPHYHAHEMIMGYAGLVV 60
	ORF47ng	MKPTKHPVKAMAPRPPYSLAALYGLSVLLWGPYGTGTHELSGPHYHAHEMIMGYAGLVV 60
	ORF47	IAFLLTAVATWTGQPPTRGVVLVGLTIFWLAARIAAPIGWGASASGILSTLFFWYGAVC 120
20	ORF47ng	IAFLLTAVATWTGQPPTRGVVLVGLTIFWLAARIAAPIGWGASASGILSTLFFWYGAVC 120
	ORF47	MALPVIRSQNRNRYVAVFALFVLGGTHAAFHVLHNGNLGGLSGLQSLVM 172
25	ORF47ng	MALPVIRSQNRNRYVAVFALFVLGGTHAAFHVLHNGNLGGLSGLQSLVMVWVGFIGLI 180

The ORF47ng nucleotide sequence <SEQ ID 711> is predicted to encode a protein comprising amino acid sequence <SEQ ID 712>:

	1 MKPTKHPVKMA MAERPPYSLA ALYGLSVLL WGPYGTGTHE LSGPHYHAHE
	51 MWSYAGLVV IAFLLTAVAT WTGQPPTRGG VINGLTAPWL AARIAAPFG
30	101 WGAASGILG LFFWYGAVC MALPVIRSQNR RRYVAVFAL FVLGGTHAAE
	151 HVQLHNGNLG GLSGLQSGLV VMVWVGFIGLI GMKIIISFPTS KRLKLPQIPS
	201 PKWVAHASLW LPMLNAILMA HRVFWLSAA FFFAAGVIPT VQVYAGGITP
	251 IETTSQGSVA GICYRLGNSS G

The predicted leader peptide and transmembrane domains are identical (except for an Ile/Ala substitution at residue 87 and an Leu/Ile substitution at position 140) to sequences in the meningococcal protein (see also *Pseudomonas stutzeri* orf396, accession number e246540):

	TM segments in ORF47ng
	INTEGRAL Likelihood = -5.63 Transmembrane 52 - 68
	INTEGRAL Likelihood = -3.88 Transmembrane 169 - 185
40	INTEGRAL Likelihood = -3.08 Transmembrane 82 - 98
	INTEGRAL Likelihood = -1.91 Transmembrane 134 - 150
	INTEGRAL Likelihood = -1.44 Transmembrane 107 - 123
	INTEGRAL Likelihood = -1.38 Transmembrane 227 - 243

Further work revealed the complete gonococcal DNA sequence <SEQ ID 713>:

45	1 ATGAAATTTA CCAACATCC CGTCTGGGCA ATGGCGTTCC GCCCGTTTTA
	51 TTCACCTGGC GCACTGTACG GCGCATTTGC CGTATTGGTG TGGGGTTTTC
	101 GCTACACGGG AACGCACGAG CTGTCGGGTT TGTATTGGCA CGCGCATGAG
	151 ATGATTGGG GTTATGCCGG TCTGCTCGTC ATCGCCTTCC TGCTGACCGC
	201 CGTGCACCAT TGGACGGGAC AGCCGCCACG GAGGGCGGCG GTTCTGCTCG
50	251 GCTTGACCGC CTTTGTGGCT GCTGCGCGGA TTGCGCGCTT TATCCCGSGT
	301 TGGGGTGGCG CGGCACGCGC CATACTCGGT ACGCTGTTTT TCTGTTAAGG
	351 CGCGGTGTGC ATGCGCTTTC CGTTATCCCG TtcgCAAAAC CGGCGCAACT
	401 ATGtcgCCGT ATTGCGCAATA TTGTGCTGCG GCGGTACGCA TCGCGcgTTC
	451 CAGGtcacAgc tGCACACGCG CACACTAGGC GACTCTTGA SCGGAITGCA
55	501 GTCGGCGCTG GTATGCGTGT CGGCGCTTAT CGGCTGATT GGATGAGA
	551 TTATTTCGTT TTTTACGTCG AACGGTGTGA ACGTGCACCA GTATCCCAAT
	601 CCGAAATGGG TGGCGCAGGC TTGCTGTGTC CTACCAATGC TGACCGGCAT

-407-

5 651 ACTGATGGCG CACGGCGTGA TGCCTTGGCT GTCGGCGGCT TCGCGGTTTG
 701 CGGCGGGCGT GATTTTACC GTACAGGTGT ACCGCTGGTG GTATAAACCC
 751 GTATTGAAG AACCGATGCT GTGGATTCTG TTTCGCGGCT ATCTGTTTAC
 801 CGGATTGGGG CTGATTGCGG TCGGCGGCTG TTAATTCAAA CCTGCCTTCC
 851 TCAATCTGGG CGTACATCTG ATCGGGGTGCG GCGGTATCGG CGTCTGACT
 901 TTGGGCATGA TGGCGGTGAC CGGCTCGGT CATAGCGCA ATTGGATTGA
 951 TCGCGCGCC AAACGGCTTC CGGTTCGGT TTGGCTGATG ATGGCGGCAA
 1001 CCGCGCTCCG TATGGTTGCC GTATTCTTCT CGGCACTCG CATACGACAC
 1051 AGCATCGCGA CGTCTTCGGT TTTGTTTGCA CTCGCGCTCG TGGTGTATCG
 1101 GTGGAATAC ATTCGGTGGC TGATCCGTCC CGGTTCCGAC GGCAGGCCCG
 1151 GTTGA

This encodes a protein having amino acid sequence <SEQ ID 714; ORF47ng-1>:

1 MKFTKHPVMA MAFRPFYSLA ALYGALSVLL WGFYGTGTHE LSGFYWHAHE
 51 MINGYAGLVV IAFLLTAVAT WTGQPPTRGG VLVLTAFAWL AARIAAFIPG
 101 WGAASGILG TLFFWYGAVC MALPVRIRSON RKNYVAVFAI FVLGGTHAAF
 151 HVQLHNGNLG GLSLGSLGSL VMVSGFIGLI GMRIISFFTS KRLNVQIPIS
 201 PKWVAQASLW LMLTAILMA HGVMPWLSAA FAFAGAVIFT VQVYRWYKYP
 251 VLKEPMLWIL FAGYLTGLG LIANGASYFK PAFALNLVHL IGVGIGVLT
 301 LGMMARTALG HTGNSIYFPF KAVPVAFWLM NATAVRMVA VFSSGTAYTH
 351 SIRTSSVLFA LALLVYAWKY IPWLIRPRSD GRPG*

ORF47ng-1 and ORF47-1 show 97.4% identity in 384 aa overlap:

		10	20	30	40	50	60
orf47-1.pep		MKFTKHPVMAAFRPFYSLAALYGALSVLLWGFYGTGTHELSGFYWHAHEMINGYAGLVV					
orf47ng-1		MKFTKHPVMAAFRPFYSLAALYGALSVLLWGFYGTGTHELSGFYWHAHEMINGYAGLVV					
		10	20	30	40	50	60
		70	80	90	100	110	120
orf47-1.pep		IAFLLTAVATWTGQPPTRGGVLVGLTI FWLAARIAAFIPGWGASASGILGTLFFWYGAVC					
orf47ng-1		IAFLLTAVATWTGQPPTRGGVLVGLTAFWLAARIAAFIPGWGAASGILGTLFFWYGAVC					
		70	80	90	100	110	120
		130	140	150	160	170	180
orf47-1.pep		MALPVRIRSONRKNYVAVFAIFVLGGTHAAFHVQLHNGNLGGLSLGSLGLVMVSGFIGLI					
orf47ng-1		MALPVRIRSONRKNYVAVFAIFVLGGTHAAFHVQLHNGNLGGLSLGSLGLVMVSGFIGLI					
		130	140	150	160	170	180
		190	200	210	220	230	240
orf47-1.pep		GTRIIISFFTSKRLNVQIPISPKWVAQASLWLPMLTAMLMHAGVLAWSAVFAFAGAVIFT					
orf47ng-1		GMRIISFFTSKRLNVQIPISPKWVAQASLWLPMLTAILMHAGVMPWLSAAFAFAGAVIFT					
		190	200	210	220	230	240
		250	260	270	280	290	300
orf47-1.pep		VQVYRWYKYPVLKEPMLWILFAGYLTGLGLIANGASYFKPAFALNLVHLIGVGIGVLT					
orf47ng-1		VQVYRWYKYPVLKEPMLWILFAGYLTGLGLIANGASYFKPAFALNLVHLIGVGIGVLT					
		250	260	270	280	290	300
		310	320	330	340	350	360
orf47-1.pep		LGMMARTALGHTGNSIYPPPKAVPVAFWLMMMAATAVRMVAVFSSGTAYTHSIRTSSVLFA					
orf47ng-1		LGMMARTALGHTGNSIYPPPKAVPVAFWLMMMAATAVRMVAVFSSGTAYTHSIRTSSVLFA					
		310	320	330	340	350	360
		370	380				
orf47-1.pep		LALLVYAWKYIPWLIRPRSDGRPGX					
orf47ng-1		LALLVYAWKYIPWLIRPRSDGRPGX					
		370	380				

Furthermore, ORF47ng-1 shows significant homology to an ORF from *Pseudomonas stutzeri*:

65 gnl|P1D|e246540 (Z73914) ORF396 protein [Pseudomonas stutzeri] Length = 396
 Score = 155 hits (389), Expect = 5e-37

Identities = 121/391 (30%), Positives = 169/391 (42%), Gaps = 21/391 (5%)

```

Query: 7  PVWAMAFRPFPYSLAALYGALSLLWGFGYTGTHELSGFY-----WHAHEMIWYGAGIV 59
          P+W +AFRPF+ +LY L++ LW +TG GF WH HEM++G+A +
5  Sbjct: 14 PIWRIAFRPFPLAGSLYALLAIPLWVAWATGLWP--GFQPTGGWIAWHRHEMLPGFAMAI 71

Query: 60  VIAFLLTAVATWTCQPPTRGGVLVGLTAFWLAARIAAFIPGWGAASGILGTLPFWYGAV 119
          V FLLTAV TWTGQ G LVGL A WLAAR+ ++ G AA L IF
10 Sbjct: 72 VAGFLLTAVQTWTCGTAFSGNRLVGLAAVWLAARL-CWLFGPLPAWLAFLDLLFLVALVM 130

Query: 120  CMALPVIKRSQNRNRYVAVFAIFVLGGTHAAFXXXXXXXXXXXXXXXXXXXXXXMSVGFGL 179
          MA + + +RNY V + ++ G +V+ + L
10 Sbjct: 131 MMAQLWAVRQKRNYPVVVLSMLGADVLIITGLLQGNDAQRQGVLAGLWLVLAALMAL 190

Query: 180  IGMRIISFFPTSKRLNVPQIPSP-KWVAQASLWPLMTAILMAHGV---MPWLSAATAFA 234
          IG R-I FFT + L P W+ A L + A+L A GV P L F A
10 Sbjct: 191 IGGRVIPFPTQRGLGKVDVAKFWWLDVALLVGTGVIALHAFGVAMRPQLLGLLGFV-A 249

Query: 235  AGVIFTVQVYRWYKPVLEPMLNIFAGYLTGLGLIAGVASFY-KPAFXXXXXXXXXXX 293
          GV +++ RW- K + K +LW L L+ + + +F A
20 Sbjct: 250 IGVGHLLRLMRWYDKGIWKVGLLWSLHVAMLMVVAAGLALWHPGLLAQSSPSLHALSV 309

Query: 294  XXXXXXXXXXXXMARTALGHTGNSIYPPKAPVPAFWLXXXXXXXXXXXXXESSGTATHSIR 353
          M+AR LGHTG + P + AF L F S +
25 Sbjct: 310 GMSGSLILAMIAVTLTGHTGRPLQLPAGIIG-AFVL---FNILGTAARVLSVAVPVGGWL 365

Query: 354  TSSVLFALALLVYAWKYPWLIRPSIDGRPG 384
          ++V + LA +Y W+Y P L+ R DG PG
30 Sbjct: 366 LAARCWTLAFALYVRYAPMLVAARVDGHPG 396

```

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 85

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 715>:

```

35      1  ..ATGCGCGTCTG AAGGTTTCAGA CGGcmTCGGT GyCGGGGAAY CAGAAAGYGGT
51      AGCGCATGCC CAATGAGACT TCGTGGGTTT TGAAGCGGGT GTTTTCCAAG
101     CGTCCCCAGT TGTGGTAACG GTATCCGGTG TCyAa:GTCA GCTTTGGYGT
151     GATGTCGAAa CGACACCGG CGATGACACC AAGACCyAmG CTGCTGATrC
201     TGTkGCTTTC GTGATAGGsa GGTTTTGTGG kmksAsyTTG TAYrATwkkG
241     CCTsCwsTG kAGmGcCKtK CkyTGGTkkA swGrwA:TAG TCGTGGTTTy
301     TktTlyyCACC GAATGAACyT GATGTTTAAC GTGTCGCTAG GCGACGCGCG
351     CGCCATATATA GGCTTTGAAT TTATCGTTGA GTTTCGAATC GTAAATGGCG
401     GACAGCCGGA GAGAGAARAC GGCGTGAAG CTGCGGTTTC CTGATGTTT
451     TGTTTGGGTT TCTTTGTAGT TCTTCTTTAT CTCTTCAGTA ACTTTTATAG
451     TAGAAGAAAT ACTTTCTTTC CATTTTCTGT AACTGGCATA ATCTGCGCGT
551     ATTCTCCAGC CGCGAAATC ..

```

This corresponds to the amino acid sequence <SEQ ID 716; ORF67>:

```

50      1  ..MFSESGDGXG XGEXEXVAHA QXDFVGFEAG VFQASPVVVVT VSGVYXQLGX
51      DVETDTGDDT KTXAADXVAF VIGRFXGXLL YXAXAXXXAX XWXXXSRGF
101     XXHRNMLNFN VSGDARADI GFEFIVEFEI VNGGQAEERN GVEAAVSLMF
151     CLGFFVVVVY LFSNFFSRRI TFFPFSVTGI ICRYSAPAII ..

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.gonorrhoeae*

ORF67 shows 51.8% identity over 199 aa overlap with a predicted ORF (ORF67ng) from

55 *N.gonorrhoeae*:

10

20

2530 **Example 86**40

45 Further work revealed the complete nucleotide sequence <SEO ID 721>:

55

```

501 GGCAGAAATG CACAGCCTGC AATCGGGTAT TTTTGTATC TTGGGTATAG
551 GTGCGACCGT TGTCGCTTGG ATTTGGTGA AAAACGCCA ACGTATCCAG
601 TTTTACCGCA GCAAATGAA AGAAAGCGG GCGCAACGCA AAGCGCCAA
651 GGCAGCCAAA AAAGCCGCGC AAGCAACA ATAA

```

- 5 This corresponds to the amino acid sequence <SEQ ID 722; ORF78-1>:

```

1 MFAPLEAFV EYGYAAFFV LVICGFGVPI PEDLTLVTVG VISGMYTNP
51 HMFAPVCMLG VLVGDCIMFA AGRIWQKIL RFKPIARIMT PKRYEQVQEK
101 FDKYGNWVLF VARFLPGLRT AVEVTAGISR KVSYLRFIM DGLAALISVP
151 IWIYLYGEYA HNIIDWLMKRN HSLQSGIFVI LGIGATVVAW IWKRRQRIG
10 FYRSKLKEKR AQRKAACA KAAQSKQ*

```

Computer analysis of this amino acid sequence predicts several transmembrane domains, and also gave the following results:

Homology with the dedA homologue of *H. influenzae* (accession number P45280)

ORF78 and the dedA homologue show 58% aa identity in 144aa overlap:

```

15 Orf78: 4 FLEAFVEYGYAAVFFVLVICGFGVPIPEDLTLVTVGVISGM--GYTNPHIMFAVGMV 61
   FL FF EYGY AV FVL+ICGFGVPIPED+TLN+GGVI+G+ N H+M V M+GV
DedA: 20 FLIGFFTEYGYWAVLFVLIICGFGVPIPEDITLVSGVIAGLYPENNVNHLMLVSMIGV 79

Orf78: 62 LVGDGIMFAAGRIWQKXILFXPIAXINTPKRYEQVQEKFDKYGNWVLFVARFLPGLRTA 121
   L GD M+ GRI+G L F PI I+T R V+EKF +YGN VLFVARFLPGLR
20 DedA: 80 LAGDCMYWLGRIYGTIKILRFPIRIVTLQRLRMVREKFSQYGNRVLFVARFLPGLRAP 139

Orf78: 122 VFVTAGISRKVSYLRFIIMDGLAA 145
   +++ +GI+R+VSY+RF++D AA
25 DedA: 140 IYVSGITRRVSYVRFVLIIDFCAA 163

```

Homology with a predicted ORF from *N. meningitidis* (strain A)

ORF78 shows 93.8% identity over a 145aa overlap with an ORF (ORF78a) from strain A of *N.*

meningitidis:

```

30 orf78.pep      10      20      30      40      50      60
   MFAPLEAFV EYGYAAVFFVLVICGFGVPIPEDLTLVTVGVISGMGYTNPHIMFAVGMV
orf78a          10      20      30      40      50      60
   MFAPLEAFV EYGYAAVFFVLVICGFGVPIPEDLTLVTVGVISGMGYTNPHIMFAVGMV

35 orf78.pep      70      80      90      100     110     120
   VLVGDGIMFAAGRIWQKXILFXPIAXINTPKRYEQVQEKFDKYGNWVLFVARFLPGLRT
orf78a          70      80      90      100     110     120
   VLVGDGIMFAAGRIWQKILFKPIARIMT PKRYAQVQEKFDKYGNWVLFVARFLPGLRT

40 orf78.pep      130     140
   AVFVTAGISRKVSYLRFIIMDGLAA
orf78a          130     140
   AVFVTAGISRKVSYLRFIIMDGLAALISVPVWIYLYGEYAHNIIDWLMKRNHSLQSGIFIA

45 orf78a          130     140     150     160     170     180

```

The complete length ORF78a nucleotide sequence <SEQ ID 723> is:

```

1 ATGTTTGCCC TTTTGAAGC CTTTTGTGCT GAATACGGCT ATGCGGCGGT
51 GTTTTTCGTT TTGCTCACT CGCGTTTCGG CCGTCCGATT CCGGAGGATT
101 TGACCTTGCT AACAGCGCGG GTGATTTCGG GTATGGGTTA TACCATCCG
151 CATATTATGT TTGCACTCGG TATGCTCGCG GTATTGGTCT GGGACGCGAT
201 CATGTTGCGC GCGGACGCA TCTGGGGGCA GAAATCCTC AGTTCACAC
251 CGATTGCGCG CATCATGACG CCGGAAOGTT ACGCACAGGT TCAGGAAAAA
301 TTGACACAAAT ACGGCAACTG GGTGTTATTT GTGCTCGTGT TCCTGCGCGG
351 TTGCGGACT GCGGTTTTCG TTACGCGCGG CATCAGCGCG AAAGTATGCT
401 ATCTGCGCTT TCTGATTATG GACGGGCTTG CCGCGCTGAT TTCGTGCCC
451 GTTTGGATTT ACTTGGGCGA GTACGGCGCG CACACATCG ATTGGCTGAT

```



```

501 GGCAGAAATG CACAGCCTGC AATCGGCGAT CTTTCATGCA TTGGCGGTGC
551 TGGCGGCGCG GCTGGCGTGG TCTGTTGGCG GCACACGCGC ACATTATCAG
601 CTTTACCGCG CACAATTGAG CGAAAACGCG GCCAAACGCA AGGCGGAAA
651 GGCAGCGAAA AAAGCGGCAC AGAAGCAGCA GTAA

```

5 This encodes a protein having amino acid sequence <SEQ ID 724>:

```

1 MFALLEAFFV EYGYAAFFV LVICGFGVPI PEDLTIVTGG VTSMGGYTNP
51 HMFVAVGMLG VLVDGIMFA AGRINGQKIL KFKPIARIMT PKRYAQVQEK
101 FDKYGNWVLF VARFLPGLRT AVFVTAGISR KVSYLRELIM DGLAALISVP
151 VMIYLYEGYA HNIDWLMAKM HSLQSGIFIA LGVLAALAW FWRKRHHYQ
10 201 LYRAQLSEKR AKRKAERKAA KAAQKQO*

```

ORF78a and ORF78-1 show 89.0% identity in 227 aa overlap:

```

10 20 30 40 50 60
orf78a.pep MFALLEAFFVEYGYAAFFVLVICGFGVPIPEDLTIVTGGVISGMGYTNPHIMFVAVGMLG
15 orf78-1 MFAFLEAFFVEYGYAAFFVLVICGFGVPIPEDLTIVTGGVISGMGYTNPHIMFVAVGMLG
10 20 30 40 50 60

70 80 90 100 110 120
orf78a.pep VLVDGIMFAAGRIWQKILKFKPIARIMTPKRYAQVQEKFDKYGNWLVFVARFLPGLRT
20 orf78-1 VLVDGIMFAAGRIWQKILKFKPIARIMTPKRYEQVQEKFDKYGNWLVFVARFLPGLRT
70 80 90 100 110 120

130 140 150 160 170 180
orf78a.pep AVFVTAGISRKVSYLRFIMDGLAALISVPVMIYLYEGYGAHNIDWLMAKMHSLSQSGIFIA
25 orf78-1 AVFVTAGISRKVSYLRFIMDGLAALISVPIYLYEGYGAHNIDWLMAKMHSLSQSGIFVI
130 140 150 160 170 180

190 200 210 220
orf78a.pep LGVLAALAWFWRKRHHYQLYRAQLSEKRKAERKAAKAAQKQX
30 orf78-1 LGIGATVVAIWWKKRQIFRSKLEKRAKRAKAAKAAQSKX
190 200 210 220

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF78 shows 97.4% identity over 38 aa overlap with a predicted ORF (ORF78ng) from *N. gonorrhoeae*:

```

40 orf78.pep XXLFXPIAXIMTPKRYEQVQEKFDKYGNWLVFVARFLPGLRTAVFVTAGISRKVSYLRF 137
orf78ng YFVLFVARFLPGLRTAVFVTAGISRKVSYLRF 32

orf78.pep IIMDGLAA 145
45 orf78ng LIMDGLAALISVPVMIYLYEGYGAHNIDWLMAKMHSLSQSGIFIALGVLAALAWFWRKR 92

```

The ORF78ng nucleotide sequence <SEQ ID 725> is predicted to encode a protein comprising amino acid sequence <SEQ ID 726>:

```

1 ..YFVLFVARFL PGLRTAVFT AGISRKVSYL RFLIMDGLAA LISVPVMIYL
51 GEYGAHNIDW LMAKMHSLSQ GFIAGLVLA AALAWFWRK RHHYQLYRAQ
50 101 LSEKRAKRAK EKAAKKAAOK QO*

```

Further work revealed the complete gonococcal nucleotide sequence <SEQ ID 727>:

```

1 atgtttgccc tttTgaagc CTTTttTTCG GAAtacgcCt atgcGGCCGT
51 GTTTTCCTGT TTTGCTCATCT GCGGTTTCGG CGTGCCGATT CCGCAAGATT
55 101 TGACCTTGGT AACGGGCGCG GTGATTTCGG GTATGGGTTA TACCAATCCG
151 CATATTATGT TGGGCTCGCG TATGCTCGCG GTGTGGCGCG GCGACGGCGT
201 GATGTTTGGC GCGGACGCGA TCTGGGGCGA GAAATCTCT AAGTTCACAC
251 CGATGSCGCG CATCATGACG CCGAAAGGTT ACGCCAGGT TACGAAAAAA
301 TTCACAAAT ACGGCAACTG GGTTCGTGTT GTCCGCCGTT TCTTCCGGGG

```

351 TTTGCGGACT GCGGTTTTCG TTACCGCCGG CATCAGCGCG AAAGTATCGT
 401 ATCTGCGGCTT TCTGATTATG GACGGGCTGG CCGCGCTGAT TTCCGTGCCG
 451 GTTTGGATT ACTTGGGCGA GTACGGCGCG CACAAATCG ATTGGCTGAT
 501 GCGCAAAATG CACAGCGTGC AATCGGGCAT CTTCATCGCA TTGGCGGTGC
 551 TGGCGGCGCG GCTGGCGTGG TTCTGGTGGC GCACAGCGCG ACATTATCAG
 601 CTTTACCGCG CACAATTGAG CGAAAAACGC GCCAACGCGA AGGGCGAAAA
 651 GGCAGCGAAA AAAGCGGCGC AGAAGCAGCA GTAA

This corresponds to the amino acid sequence <SEQ ID 728; ORF78ng-1>:

1 MFALLEAFFV EYGYAAFFV LVICGFGVPI PEDTLVTGG VISGMGYTNP
 51 HIMEFAVGMGLG VLADGGMFA AGRIWGQKIL KFKPIARINT PKRYAQVOEK
 101 FDKYGNWVLF VARFLPGLRT AVFVTAGISR KVSYLRFILM DGLAALISVP
 151 VWIYLGEYGA HNIDWLMAKM HSLQSGIFIA LGVLAALAW FWRKRHHYQ
 201 LYRAQLSEKR AKRKAERKAA KAAQKQ*

ORF78ng-1 and ORF78-1 show 88.1% identity in 227 aa overlap:

15	orf78-1.pep	10	20	30	40	50	60
	MFALLEAFFVEYGYAAFFVLVICGFGVPIPEDTLVTGGVISGMGYTNP						
	orf78ng-1	10	20	30	40	50	60
	MFALLEAFFVEYGYAAFFVLVICGFGVPIPEDTLVTGGVISGMGYTNP						
20	orf78-1.pep	70	80	90	100	110	120
	VLVGDGMFAAGRIWGQKILKFKPIARINTPKRYAQVOEKFDKYGNWVLFVARFLPGLRT						
	orf78ng-1	70	80	90	100	110	120
	VLVGDGMFAAGRIWGQKILKFKPIARINTPKRYAQVOEKFDKYGNWVLFVARFLPGLRT						
25	orf78-1.pep	130	140	150	160	170	180
	AVFVTAGISRKVSYLRFIMDGLAALISVPVWVWYLYGEYGAHNIDWLMAKMHSLSQSGIFIA						
	orf78ng-1	130	140	150	160	170	180
	AVFVTAGISRKVSYLRFIMDGLAALISVPVWVWYLYGEYGAHNIDWLMAKMHSLSQSGIFIA						
30	orf78-1.pep	190	200	210	220		
	LGIGATVVMVWKKRQRIQFYRSKLKRAQKAAKAAQKQX						
	orf78ng-1	190	200	210	220		
	LGVLAAALANFWRKRHHYQLYRAQLSEKRAKRAKAAKAAQKQX						

Furthermore, orf78ng-1 shows homology to the dedA protein from *H. influenzae*:

40 sp|P45280|YG29_HAEIN_HYPOTHETICAL_PROTEIN_HIL629 >gi|1073983|pir||D64133 dedA
 protein (dedA) homolog - Haemophilus influenzae (strain Rd KW20)
 >gi|1574476 (U32836) dedA protein (dedA) [Haemophilus influenzae] Length = 212
 Score = 223 bits (563), Expect = 7e-58
 Identities = 108/182 (59%), Positives = 140/182 (76%), Gaps = 2/182 (1%)

45 Query: 5 LEAFFVEYGYAAFFVLVICGFGVPIPEDTLVTGGVISGM--GYTNP HIMEFAVGMGLV 62
 L EF EYGY AV FVLICGFGVPIED+TLV+GGVI+G N H+M V M+GVL
 Sbjct: 21 LIGFFTEYGYWAVLFVLICGFGVPIPEDTLVTGGVIAGIYEPVNVSHIMLVSMIGVL 80

50 Query: 63 AGDGMFAAGRIWGQKILKFKPIARINTPKRYAQVOEKFDKYGNWVLFVARFLPGLRTAV 122
 AGD M+ GRI+G KIL+P+PI RI+T +R V+EK+YGN VLFVARFLPGLR +
 Sbjct: 81 AGDSCMYLGRIGYTKILRFRIPIRIVTLQRLRMVREKFSQYGNRVLVVARFLPGLRPI 140

55 Query: 123 FVTAGISRKVSYLRFIMDGLAALISVPVWVWYLYGEYGAHNIDWLMAKMHSLSQSGIFIALG 182
 ++ +GI+R+VS+YR+P++D AA+ISVP+WYLYGE GA N+DNL ++ Q I+I +G
 Sbjct: 141 YWVGIGITRRVSYVRFLVDFCAALISVPVWVWYLYGELGAKNLDLWLTQIQKQGVYLYFTIG 200

60 Query: 183 VL 184
 L
 Sbjct: 201 YL 202

Based on this analysis, including the presence of putative transmembrane domains, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 87

- 5 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 729>:

```

1 ATGAAAAAAT TATTGGCGGC CGTGATGATG GCAGGTTTGG CAGGCGCGGT
51 TTCCGCGCGC GGAGTCCACG TTGAGGACGG CTGGGCGCGC ACCACGTCG
101 AAGGTATGAA AATAGGCGGC GCGTTTCATGA AAATCCACAA CGACGAAGCC
151 AAACAAGACT TTTTGCTCGG CGGAAGCAGC CCGGTTGCGG ACCGCGTCGA
201 AGTCATACCC CACATCAACG ACACGCGGCT GATCGGGATG CGCGAAGT CG
251 AAGCGGCGCT GCCTTTGGAA CGCAATCCG TTACCGAATC CAACCCCGGC
301 AGCTATCATG TGATCTTTAT GGGTTTGAAA AAACAATTAA AAGAGGCGCA
351 TAAAAATCCC GTTACCCTGA AATTAAAAAA CGCCAAAGCG CAACCGGTCC
401 AACTGGAAGT CAAATCGCGC CCGATGCCGG CAATGAACCA C...
```

- 15 This corresponds to the amino acid sequence <SEQ ID 730; ORF79>:

```

1 MKKLLAAVMM AGLAGAVSAA GVHVEDGWAR TTVEGMKIGG AFMKIHDEA
51 KQDFLLGGSS PVADRVVHT HINDNGVMRM REVEGGVPLE AKSVTELKPG
101 SYHVMFMGLK KQLKEGDKIP VTLKFNKA QTVQLEVKIA PMPAMNH...
```

Further work revealed the complete nucleotide sequence <SEQ ID 731>:

```

20 1 ATGAAAAAAT TATTGGCGGC CGTGATGATG GCAGGTTTGG CAGGCGCGGT
51 TTCCGCGCGC GGAGTCCACG TTGAGGACGG CTGGGCGCGC ACCACGTCG
101 AAGGTATGAA AATAGGCGGC GCGTTTCATGA AAATCCACAA CGACGAAGCC
151 AAACAAGACT TTTTGCTCGG CGGAAGCAGC CCGGTTGCGG ACCGCGTCGA
201 AGTCATACCC CACATCAACG ACACGCGGCT GATCGGGATG CGCGAAGT CG
251 AAGCGGCGCT GCCTTTGGAA CGCAATCCG TTACCGAATC CAACCCCGGC
301 AGCTATCATG TGATCTTTAT GGGTTTGAAA AAACAATTAA AAGAGGCGCA
351 TAAAAATCCC GTTACCCTGA AATTAAAAAA CGCCAAAGCG CAACCGGTCC
401 AACTGGAAGT CAAATCGCGC CCGATGCCGG CAATGAACCA CGGTATCATC
451 CACGGCGAAG CGCATCAGCA CTA
```

- 30 This corresponds to the amino acid sequence <SEQ ID 732; ORF79-1>:

```

1 MKKLLAAVMM AGLAGAVSAA GVHVEDGWAR TTVEGMKIGG AFMKIHDEA
51 KQDFLLGGSS PVADRVVHT HINDNGVMRM REVEGGVPLE AKSVTELKPG
101 SYHVMFMGLK KQLKEGDKIP VTLKFNKA QTVQLEVKIA PMPAMNHGHI
151 HGEAHQH*
```

- 35 Computer analysis of this amino acid sequence revealed a putative leader peptide and also gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF79 shows 94.6% identity over a 147aa overlap with an ORF (ORF79a) from strain A of *N.meningitidis*:

```

40      10      20      30      40      50      60
orf79.pep MKKLLAAVMMAGLAGAVSAAAGVHVEDGWAR TTVEGMKIGGAFMKIHNDKQDFLLGGSS
orf79a     MKKLLAAVMMAGLAGAVSAAAGIHVEDGWAR TTVEGMKMGGA FMKIHNDKQDFLLGGSS
      10      20      30      40      50      60
45      70      80      90     100     110     120
orf79.pep PVADRVVHTHINDNGVMRMREVEGGVPLEAKSVTELKPGSYHVMFMGLKKQLKEGDKIP
orf79a     FVADRVVHTHINDNGVMRMREVEGGVPLEAKSVTELKPGSYHVMFMGLKKQLKEGDKIP
50      70      80      90     100     110     120
```

-414-

```

              130      140
orf79.pep    VTLKFKNAKAQTVQLEVKIAPMPAMNH
              |||||
5   orf79a    VTLKFKNAKAQTVQLEVKIAPMSAMDHGHHGEAHQHX
              130      140      150

```

The complete length ORF79a nucleotide sequence <SEQ ID 733> is:

```

1   ATGAAANAAC TATTGGCAGC CGTGATGATG GCAGGTTTGG CAGGCGGGGT
51  TTCGCGCGCC GGATCCACG TTAGGACGGC CTGGGCGCGC ACCACCGTCG
101 AAGGTATGAA AATGGCGCGG GCGTTTCATGA AATCCACAA CGACGAAGCC
151 AAACAAGACT TTTGCTCGG CGGAGCAGC CCGTTTGGCG ACCGCGTCGA
201 AGTGCAATACC CATATCAATG ATACCGTGT GATGCGGATG CGCGAAGTCG
251 AAGCGCGCGT GCGTTGGAG GCGAATCCG TTACCGGAAT CAACCCGGCG
301 AGCTATCATG TCATGTTTAT GGGTNTGAAA AAACAATTAA AAGANGGCGA
15 351 CAAGATTCCC GTTACCCTGA AATTTAAAA CGCCAAAGCA CAACCCGTCC
401 AACTGGAGT CAAACCGCG CCGATGTCGG CAATGGACCA CGGTCATCAC
451 CACGCGAAG CGCATCAGCA CTA

```

This encodes a protein having amino acid sequence <SEQ ID 734>:

```

1   MKLLAAVMM AGLAGAVSA GIHVEDGWAR TTVEGMGGG AFMKIHND EA
20 51  RDPFLGGSS PVADRVVHT HINDNGVMRM REVEGVPLE AKSVTELKPG
101 SYHVMFMGLK KQLXGDKIP VTLKFKNAKA QTVQLEVKTA PMSAMDHGH
151 HGEAHQH*

```

ORF79a and ORF79-1 show 94.9% identity in 157 aa overlap:

```

25   orf79a.pep    MKLLAAVMMAGLAGAVSAAGIHVEDGWAR TTVEGMGGGAFMKIHND EAKQDFLLGGSS
              || |||||
orf79-1    MKLLAAVMMAGLAGAVSAAGVHVEDGWAR TTVEGMKIGGAFMKIHND EAKQDFLLGGSS
              10      20      30      40      50      60
30   orf79a.pep    PVADRVVHTHINDNGVMRMREVEGGVPLEAKSVTELKPGSYHVMFMGLK KQLXGDKIP
              |||||
orf79-1    PVADRVVHTHINDNGVMRMREVEGGVPLEAKSVTELKPGSYHVMFMGLK KQLXGDKIP
              70      80      90      100     110     120
35   orf79a.pep    VTLKFKNAKAQTVQLEVKTA PMSAMDHGHHGEAHQHX
              |||||
orf79-1    VTLKFKNAKAQTVQLEVKIAPMPAMNHGHHGEAHQHX
              130     140     150
40   orf79a.pep    VTLKFKNAKAQTVQLEVKTA PMSAMDHGHHGEAHQHX
              |||||
orf79-1    VTLKFKNAKAQTVQLEVKIAPMPAMNHGHHGEAHQHX
              130     140     150

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF79 shows 96.1% identity over 76 aa overlap with a predicted ORF (ORF79ng) from *N.gonorrhoeae*:

```

45   orf79.pep    FMKIHND EAKQDFLLGGSSPVADRVVHTHINDNGVMRMREVEGGVPLEAKSVTELKPGS 101
              |||||
orf79ng          INDNGVMRMREVEGGVPLEAKSVTELKPGS 30
50   orf79.pep    YHVMFMGLKKQLKEGDKIPVTLKFKNAKAQTVQLEVKIAPMPAMNH 147
              |||||
orf79ng          YHVMFMGLKKQLKEGDKIPVTLKFKNAKAQTVQLEVKTA PMSAMNHGHHGEAHQH 86

```

An ORF79ng nucleotide sequence <SEQ ID 735> was predicted to encode a protein comprising amino acid sequence <SEQ ID 736>:

```

1   ..INDNGVMRM EVKGVPLEA KSVTELKPGS YHVMFMGLKK QLKEGDKIPV
55 51  TLKFKNAKAQ TVQLEVKTA PMSAMNHGHH GEAHQH*

```

Further work revealed the complete gonococcal DNA sequence <SEQ ID 737>:

```

      1 ATGAAAAAAT TATTGGCAGC CGTGATGATG GCAGGTTTGG CAGGCGCGGT
    51 TTccgcgcgc GgagTccAtG TCGAggACGG CTGGGCGCGC accaCTGtcg
   101 aaggtATgaa aatggGCGGC GcTtCATga aaATCCACAA CGACGaaGcc
   151 atacaagACT tttgtcTcgg CGGaaagcatg cccgttgcg accgcGTCGA
   201 AGTGCAatca cacATCAAGC ACRAAGCGGT GATGCGTATG CGCGAAGTCA
   251 AAGGCGCGGT GCCTTTGGAG GCGAAGTCTG TACCGAATC CAACCCCGCG
   301 AGCTATCAAG TGATGTTTAT GGGTTTAAAA AAAACACTGA AAGAGCGGCA
   351 CAGATTCCCT CTCACCTGTA RAATTAAAAA CGCCAAAGCG CAACCGTCC
   401 AACTGGAAGT CAAAACCGCG CGATGTCGCG CAATGAACCA CGGTCAATCA
  10 451 CACGGCGAAG CGCATCAGCA CTTAA

```

This corresponds to the amino acid sequence <SEQ ID 738; ORF79ng-1>:

```

      1 MKKLLAAVMM AGLAGAVSAA GVHVEDGWAR TTVGMMKMG AFMKIHNDDEA
    51 IQDFVLGGSM PVADRVVHT HINDNGVMRM REVKGVPLE AKSVTELPKG
   101 SYHVMFMGLK KQLKEGDKIP VTLKFKNKA QTVQLEVKTA PMSAMNHGH
   151 HGEAHQH*

```

ORF79ng-1 and ORF79-1 show 95.5% identity in 157 aa overlap:

```

      10      20      30      40      50      60
  orf79-1.pep MKKLLAAVMMAGLAGAVSAAAGVHVEDGWARTTVEGMKMGAFMKIHNDDEAKQDFLLGGSS
      |||
  20 orf79ng-1 MKKLLAAVMMAGLAGAVSAAAGVHVEDGWARTTVEGMKMGAFMKIHNDDEAIQDFVLGGSM
      |||
      10      20      30      40      50      60
      70      80      90      100     110     120
  25 orf79-1.pep PVADRVVHTHINDNGVMRMREVKGVPLEAKSVTELPKGSYHVMFMGLKKQLKEGDKIP
      |||
  orf79ng-1 PVADRVVHTHINDNGVMRMREVKGVPLEAKSVTELPKGSYHVMFMGLKKQLKEGDKIP
      70      80      90      100     110     120
      130     140     150
  30 orf79-1.pep VTLKFKNKAQTVQLEVVKIAPMAMNHGHGHGEAHQH
      |||
  orf79ng-1 VTLKFKNKAQTVQLEVKTAPMSAMNHGHGHGEAHQH
      130     140     150

```

Furthermore, ORF79ng-1 shows significant homology to a protein from *Aquifex aeolicus*:

```

  35 gi|2983695 (AE000731) putative protein [Aquifex aeolicus] Length = 151
      Score = 63.6 bits (152), Expect = 6e-10
      Identities = 38/114 (33%), Positives = 58/114 (50%), Gaps = 1/114 (0%)

  40 Query: 24 VEDGWARTTVEGMKMGAFMKIHNDDEAIQDFVLGGSMFPVADRVVHTHINDNGVMRMREV 83
      V+ W G M I N+ D++G +A RVE+H ++N V +M
      Sbjct: 27 VKHPVWMEPPPGPNTTMMGMIIVNEGDEPDYLIKATDIAQVELHKTVTIENDVAKMVPQ 86

  45 Query: 84 KGGVPLEAKSVTELPKGSYHVMFMGLKKQLKEGDKIPVTLKFKNKAQTVQLEV 137
      V+ + + K E K YHVM +GLKK+-KEGDK+ V L F+ + TV+ V
      Sbjct: 87 ER-IEIPPGKVEFKHHGYHVMIIGLKKRIKEGDKVKEVLEIFKSGKITVEAFV 139

```

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF79-1 (15.6kDa) was cloned in the pET vector and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 18A shows the results of affinity purification of the His-fusion protein. Purified His-fusion protein was used to immunise mice, whose sera were used for ELISA (positive result) and FACS analysis (Figure 18B) These experiments confirm that ORF79-1 is a surface-exposed protein, and that it is a useful immunogen.

Example 88

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID 739>:

```

1   ATGACGGTAA CTGCGGCCGA AGCGGGCAAA GCTGCCAAGG CGTTAAAAAA
51  ATATCTGATT ACGGGCATT TGGCTGGCT GCCGATTGCG GTACGCGTTT
101 GGGTGGTTTC CTATATCGTT TCCGCGTCCG ATCAGCTCGT CAACCTGCTG
151 CCGAAGCAAT GCGCGCCGCA ATATGTTTGT GGGTTTAATA TCCCGGGGCT
201 GGGCGTTATC GTTGCCATTG CCGTATTGTT TGTAACCGGA TTGTTTGGCG
251 CCAACGTATT GGGTCGGCAG ATCCTCGCCG CGTGGGACAG CCGTGGGGGG
10 301 CGGATTCCGG TTGTGAATC CATCTATTCG AGTGTGAAA AAGTATCCGA
351 ATAGCTGCTG TCGACAGCA GCGCTCGTGT TAAACGCCG GTACTCGTGC
401 CGTTTCCCCA GCCCGGTATT TGGACGATyG CTTTCGTGTC AGGCGAGGTG
451 TCGAATGCGG TTAAGGCCGC ATTGCGGAA GACGCGGATT ATCTTTCCGT
501 GTATGTTCCG ACCACGCCGA ATCCGACCGG CGGTTACTAT ATTATGGTAA
15 551 AGAAAAGCGA TGTGCGCGAA CTCGATATGA CGGTGGACGA AaCATTGAAA
601 TATGTGATTT CGCTGGGTAT GGTATCCCT GACGACCTGC CCGTCAAAAC
651 ATTGGCAsGA CCTATGCCGT CTGAAAAGGC GGATTTGCCC GAACACCAAT
701 AA

```

This corresponds to the amino acid sequence <SEQ ID 740; ORF98>:

```

20 1   MVTVAEEGK AAKALKKYL IGLVWLPIA VIVVWVSIV SASDQLVNL
51  PKQWRPQYVL GFNIPGLVI VAIIVLVFTG LFAANVLGRQ ILAANDSLLG
101 RIFVVKSIYS SVKKVSEIWL SSSRSFKTP VLVFFPQGI WTIAFVSGQV
151 SNAVKALFK DGYLSVIVP TPNFTGGYI IMVKSDVRE LMSVDLXK
201 YVLSLGMVIP DDLFVKTLAK FMPSEKADLF EQQ*

```

Further work revealed the complete nucleotide sequence <SEQ ID 741>:

```

1   ATGACGGTAA CTGCGGCCGA AGCGGGCAAA GCTGCCAAGG CGTTAAAAAA
51  ATATCTGATT ACGGGCATT TGGCTGGCT GCCGATTGCG GTACGCGTTT
101 GGGTGGTTTC CTATATCGTT TCCGCGTCCG ATCAGCTCGT CAACCTGCTG
151 CCGAAGCAAT GCGCGCCGCA ATATGTTTGT GGGTTTAATA TCCCGGGGCT
201 GGGCGTTATC GTTGCCATTG CCGTATTGTT TGTAACCGGA TTGTTTGGCG
251 CCAACGTATT GGGTCGGCAG ATCCTCGCCG CGTGGGACAG CCGTGGGGGG
30 301 CGGATTCCGG TTGTGAATC CATCTATTCG AGTGTGAAA AAGTATCCGA
351 ATAGCTGCTG TCGACAGCA GCGCTCGTGT TAAACGCCG GTACTCGTGC
401 CGTTTCCCCA GCCCGGTATT TGGACGATG CTTTCGTGTC AGGCGAGGTG
35 451 TCGAATGCGG TTAAGGCCGC ATTGCGGAA GACGCGGATT ATCTTTCCGT
501 GTATGTTCCG ACCACGCCGA ATCCGACCGG CGGTTACTAT ATTATGGTAA
551 AGAAAAGCGA TGTGCGCGAA CTCGATATGA CGGTGGACGA AaCATTGAAA
601 TATGTGATTT CGCTGGGTAT GGTATCCCT GACGACCTGC CCGTCAAAAC
651 ATTGGCAGGA CCTATGCCGT CTGAAAAGGC GGATTTGCCC GAACACCAAT
40 701 AA

```

This corresponds to the amino acid sequence <SEQ ID 742; ORF98-1>:

```

1   MTEAAEEGK AAKALKKYL IGLVWLPIA VIVVWVSIV SASDQLVNL
51  PKQWRPQYVL GFNIPGLVI VAIIVLVFTG LFAANVLGRQ ILAANDSLLG
101 RIFVVKSIYS SVKKVSEIWL SSSRSFKTP VLVFFPQGI WTIAFVSGQV
151 SNAVKALFK DGYLSVIVP TPNFTGGYI IMVKSDVRE LMSVDLXK
45 201 YVLSLGMVIP DDLFVKTLAK FMPSEKADLF EQQ*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF98 shows 96.1% identity over a 233aa overlap with an ORF (ORF98a) from strain A of *N.meningitidis*:

```

50 orf98.pep      10      20      30      40      50      60
orf98a      MVTVAEEGKAAKALKKYLITGILVWLPIAVIVVWVSIVSASDQLVNLPKQWRPQYVL
|| |||||
55 orf98a      MTEPAEEGKAAKALKKYLITGILVWLPIAVIVVWVSIVSASDQLVNLPKQWRPQYVL
      10      20      30      40      50      60

```

-417-

		70	80	90	100	110	120
5	orf98.pep	GFNIPGLGVIAIAVLFTVGLFAANVLGRQILAAWDSLLGRIPVVKSIYSSVKKVSEYVL					
	orf98a	GFNIPGLGVIAIAVLFTVGLFAANVLGRQILAAWDSLLGRIPVVKSIYSSVKKVSEYVL					
		70	80	90	100	110	120
10	orf98.pep	SDSSRSFKTPVLVFPFQPGIWTIAFVSQVSNVAKAALPKDGDYLSVYVPTTNPPTGGYY					
	orf98a	SDSSRSFKTPVLVFPFQPGIWTIAFVSQVSNVAKAALPKDGDYLSVYVPTTNPPTGGYY					
		130	140	150	160	170	180
15	orf98.pep	IMVKKSDVRELDMSVDEALKYVISLGMVIPDDLFPVKTLAGMPSEKADLPEQQX					
	orf98a	IMVKKSDVRELDMSVDEALKYVISLGMVIPDDLFPVKTLAGMPSEKADLPEQQX					
		190	200	210	220	230	
		190	200	210	220	230	

The complete length ORF98a nucleotide sequence <SEQ ID 743> is:

20	1	ATGACGGAAC	CTGCGGCCGA	AGGCGGCAAA	GCTGCCACAG	CGTTAAAAAA
	51	ATATCTGATT	ACGGCGCATT	TGGTCTGGCT	GCGGATTCCG	GTAACGTTT
	101	GGGTGCTTC	CTATATCGTT	TCCGCTCGCT	ATCAGCTCGT	CAACTGCTG
	151	CCGAAGCAAT	GGCGGCCGCA	ATATGTTTGT	GGGTTTAATA	TCCCGGGCT
	201	GGCGCTTATC	GTTGCCATCG	CCGTATTGTT	TGTAACCGGA	TTATTGTCG
25	251	CAAAAGTATT	GGCGGCCGCA	ATTCTTGCGC	CGTGGACAG	CTGTTGGGG
	301	CGGATTCGG	TTGTGAAGTC	CATCTATTCT	AGTGTGAAA	AAGTATCCGA
	351	NTCGTTGCTG	TCCGACAGCA	GCGCTTCGTT	TAAACACCA	GTACTCGTGC
	401	CGTTTCCCA	ATCGGCTATT	TGGCAATCG	CATTGCTGTC	CGGTCAAGTG
	451	TGCAATCGG	TTAAGGCGCG	ATTGCGAAG	GACGCGGATT	ATCTTTCGTT
30	501	GTATGTTCCG	ACCAACCGCA	ATCCGACCG	CGGTACTACT	ATTATGGTAA
	551	AGAAAAGCGA	TGTGCGCGAA	CTGATATGA	CGGTGACCA	ACGCTGAAA
	601	TATGTGATTT	CGTGGGATAT	GCTCATCCCT	GACGACCTGC	CGGTCAAAAC
	651	ATTGGCAGGA	CCTATGCCCT	CTGAAAAGGC	GGATTTGCCG	GACACACAT
	701	AA				

35 This encodes a protein having amino acid sequence <SEQ ID 744>:

	1	MTEPAAEGGK	AAKALKKYL	ITGILVWLPIA	VTWVVSIV	SASDQLVNL
	51	PKQWRPQYVL	GFNIPGLGVI	VAIAVLFTVGL	LFAANVLGRQ	ILAAWDSLLG
	101	RIPVVKSIYS	SVKKVXSLL	SDSSRSFKTP	VLVFPFQSGI	WTIAFVSQV
40	151	SNVAKAALPK	DGDYLSVYVP	TTPNPPTGGY	IMVKKSDVRE	LDMSVDEALK
	201	YVISLGMVIP	DDLFPVKTLAG	MPSEKADLE	EQQ*	

ORF98a and ORF98-1 show 98.7% identity in 233 aa overlap:

		10	20	30	40	50	60
45	orf98a.pep	MTEPAAEGGKA	AAKALKKYLIT	GILVWLPIAV	TWVVSIVYSAS	DQLVNL	PKQWRPQYVL
	orf98-1	MTEPAAEGGKA	AAKALKKYLIT	GILVWLPIAV	TWVVSIVYSAS	DQLVNL	PKQWRPQYVL
		10	20	30	40	50	60
50	orf98a.pep	GFNIPGLGVIAIAVLFTVGLFAANVLGRQILAAWDSLLGRIPVVKSIYSSVKKVSEYVL					
	orf98-1	GFNIPGLGVIAIAVLFTVGLFAANVLGRQILAAWDSLLGRIPVVKSIYSSVKKVSEYVL					
		70	80	90	100	110	120
55	orf98a.pep	SDSSRSFKTPVLVFPFQPGIWTIAFVSQVSNVAKAALPKDGDYLSVYVPTTNPPTGGYY					
	orf98-1	SDSSRSFKTPVLVFPFQPGIWTIAFVSQVSNVAKAALPKDGDYLSVYVPTTNPPTGGYY					
		130	140	150	160	170	180
60	orf98a.pep	IMVKKSDVRELDMSVDEALKYVISLGMVIPDDLFPVKTLAGMPSEKADLPEQQX					
	orf98-1	IMVKKSDVRELDMSVDEALKYVISLGMVIPDDLFPVKTLAGMPSEKADLPEQQX					
		190	200	210	220	230	
		190	200	210	220	230	

Homology with a predicted ORF from *N.gonorrhoeae*

ORF98 shows 95.3% identity over a 233 aa overlap with a predicted ORF (ORF98ng) from *N.gonorrhoeae*:

		10	20	30	40	50	60	
5	orf98.pep	MTVTAEEGGKAALKKYLITGILVWLP	IAVTWVWVSIVSASDQLVNL	LPKQWRPQYVL				60
	orf98ng	MTEPAEEGGKAALKKYLITGILVWLP	IAVTWVWVSIVSASDQLVNL	LPKQWRPQYVL				60
10	orf98.pep	GFNIPGLGVIAIAVLFTGLFAANVLGRQILA	AWDSLLGRIPVVKSIYSSVKVSEYVL					120
	orf98ng	GFNIPGLGVIAIAVLFTGLFAANVLGRQILA	AWDSLLGRIPVVKSIYSSVKVSESL					120
15	orf98.pep	SDSSRSFKTPVLVFPFQPGIWTIAFVSGQVS	NAVKAALPXDGDYLSVYVPTTNP	TGGYY				180
	orf98ng	SDSSRSFKTPVLVFPFQPGIWTIAFVSGQVS	NAVKAALPQDGDYLSVYVPTTNP	TGGYY				180
20	orf98.pep	IMVKKSDVRELDMSVDEALKYVILGMVIP	PDDLVPKTLAXMPEKADLPEQQ					233
	orf98ng	IMVKKSDVRELDMSVDEALKYVILGMVIP	PDDLVPKTLAXMPEKADLPEQQ					233

20 The complete length ORF98ng nucleotide sequence <SEQ ID 745> is predicted to encode a protein having amino acid sequence <SEQ ID 746>:

1	MTEPAEEGGK	AAKALKKYL	TGILVWLP	IAVTWVWVSIV	SASDQLVNL	
51	PKQWRPQYVL	GFNIPGLGVI	VAIAVLFTG	LFAANVLGRQ	ILAAWDSLLX	
101	RIPVVKSIYS	SVKVSSESL	SDSSRSFKTP	VLVFPQSGI	WTIAFVSGQV	
151	SNVKAALPQ	DGDYLSVYVP	TTPNPTGGYY	IMVKKSDVRE	LDMSVDEALK	
201	YVISLGMVIP	DDLVPKTL	AGMPEKADL	PEQQ		

Further work revealed the complete nucleotide sequence <SEQ ID 747>:

1	ATGACGGAAC	CTGCGGCCGA	AGCGGCCAAA	GCTGCCAAGG	CGTTAAAAAA	
51	ATATCTGATT	ACAGGCAATT	TGCTTGGCT	CCGATTTGGT	GTAACGGTTT	
101	GGGTGGTTTC	CTATATCGTT	TCCGCGTCCG	ACCAGCTTGT	CAACCTGCTG	
151	CCGAAGCAAT	GGCGGCCGCA	ATATGTTTGT	GGGTTAATA	TCCCCGGGCT	
201	CGCGCTTATT	GTTCGCATTG	CCGTATGTTT	TGTAACCGGA	TTATTTCGCG	
251	CAAAAGTGT	GGCGGCCGAG	ATTCTTGC	CGTGGGACAG	CCTGTTgggg	
301	cggatTCGCG	TTGTCAATC	CATCTATTCG	AGTGTGAAAA	AAGTATCCGA	
351	ATCGCTGCTG	TCCGACAGCA	GCCGTTGTT	TAAAACCGCG	GTACTCGTGC	
401	CGTTTCCCCA	ATCGGGTATT	TGGCAATTC	CATTCTGTCT	CGGTCAAGTG	
451	TGCAATGGGG	TAAAGCCGCG	ATTGCCGACG	GATGGCGATT	ATCTTTCGCT	
501	GATATGCTCC	ACCAACGCCA	ACCCGACCGG	CGGTACTACT	ATTATGGTAA	
551	AGAAAAGCGA	TGTGCGCGAA	CTCGATATGA	CGGTGACAGA	AGCGTTGAAA	
601	TATGTGATTT	CGCTGGGTAT	GTCATCCCT	GACGACCTCG	CGCTCAAAAC	
651	ATTGGCAGGA	CCTATGCCGC	CTGAAAAGGC	GGAGTTGCCG	GAACAACAA	
701	AA					

This corresponds to the amino acid sequence <SEQ ID 748; ORF98ng-1>:

1	MTEPAEEGGK	AAKALKKYL	TGILVWLP	IAVTWVWVSIV	SASDQLVNL	
51	PKQWRPQYVL	GFNIPGLGVI	VAIAVLFTG	LFAANVLGRQ	ILAAWDSLLG	
101	RIPVVKSIYS	SVKVSSESL	SDSSRSFKTP	VLVFPQSGI	WTIAFVSGQV	
151	SNVKAALPQ	DGDYLSVYVP	TTPNPTGGYY	IMVKKSDVRE	LDMSVDEALK	
201	YVISLGMVIP	DDLVPKTL	AGMPEKADL	PEQQ		

ORF98ng-1 and ORF98-1 show 97.9% identity in 233 aa overlap:

		10	20	30	40	50	60	
50	orf98-1.pep	MTEPAEEGGKAALKKYLITGILVWLP	IAVTWVWVSIVSASDQLVNL	LPKQWRPQYVL				
	orf98ng-1	MTEPAEEGGKAALKKYLITGILVWLP	IAVTWVWVSIVSASDQLVNL	LPKQWRPQYVL				
55		10	20	30	40	50	60	
	orf98-1.pep	GFNIPGLGVIAIAVLFTGLFAANVLGRQILA	AWDSLLGRIPVVKSIYSSVKVSESL					
		70	80	90	100	110	120	
		GFNIPGLGVIAIAVLFTGLFAANVLGRQILA	AWDSLLGRIPVVKSIYSSVKVSESL					

5	orf98ng-1	GFNIPGLGVIVAIIVLFTGLFAANVLGRQILAAWDSLGRIPVVKSIYSSVVKVSESLL	70	80	90	100	110	120
			130	140	150	160	170	180
	orf98-1.pcp	SDSSRSFKTPVLVFPFQPGIWTIAFVSQVSNVAKAALPKDGDYLSVYPTTTFNETGGY						
10	orf98ng-1	SDSSRSFKTPVLVFPFQPGIWTIAFVSQVSNVAKAALPKDGDYLSVYPTTTFNETGGY	130	140	150	160	170	180
			190	200	210	220	230	
	orf98-1.pcp	IMVKKSDVRELDMSVDALKYIISLGMVIPDDLVPKTLAGMPSEKADLPEQQX						
15	orf98ng-1	IMVKKSDVRELDMSVDALKYIISLGMVIPDDLVPKTLAGMPSEKADLPEQQX	190	200	210	220	230	
			190	200	210	220	230	
	orf98-1.pcp	IMVKKSDVRELDMSVDALKYIISLGMVIPDDLVPKTLAGMPSEKADLPEQQX						

Based on this analysis, including the fact that the putative transmembrane domains in the gonococcal protein are identical to the sequences in the meningococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 89

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 749>:

1	ATGAAAACGG	TAGTCTGGAT	TGTCGCTCTG	TTTGCCGCGC	CGCTCGGACT
51	GGCGCTGGCT	TCGGGCATT	ACACCGGCGA	CGTGATATAC	GTACTCGGAC
101	AGACCATGCT	CAGAATCAAC	CTGCACGCT	TTGTGTTAGG	TTGCTGATT
151	GGCGCTGGCT	TGTGGTATT	CTTGTTTAAA	TTTATTATCG	GgGgTACTCA
201	ATATCCCCGA	AAAGATGCAG	CGTTTCGGTT	CGGChCGTAA	AGGCKCAAG
251	ssCGsGCTTG	CCTTGACAA	GGCGGGTTG	GGCTATTTTG	AAGGCGGTT
301	TGAAAGGCG	GAACTAGAAG	CCTCAGCGT	GTGGTCAAC	AAAGTAGGCG
351	GaGAGACAAC	CGGACTTTG	CATTGATCT	GtGCGCGCAC	GCGCGCGGAC
401	AGATGGAAAA	CATCGAsTG	CGCGACGTT	ATCTTGCGGA	AATGCGCAAA
451	CTGCCGAAAA	AACAGCAGCT	TTCCCGTTAT	CTTTTGTGTG	CGGAATCGCG
501	GTGAAACCG	CGCGATTACG	AAGCGCGGGA	AGCCAACTCT	CATGCGCGGG
551	OGAAGATGAA	TGCCAACCTT	ACGCGCTCG	TGCGTCTGCA	ATTGTTTAC
601	GCTTTGACAA	GGGCGGAG	GTCGACGTT	CTGGCAAAA	CGAAAAACT
651	TTCCAAAGCG	GGCGGCTTG	GCAATCGGA	AATGGAACGG	TATCAAAATT
701	GGCATATCTG	GTCGCCAGCT	GGCGATGCT	GCGGATGCGC	CGCGTTGAA
751	AACCTGCGCTG	AAGCGGATTC	CGCAGCGCT	CAAAAACGGG	GAATTGAGCG
801	TATCGGTTGC	GGRAAAGTAC	GAACGTTTGG	GACTGTATGC	CGATCGCGTC
851	AAATGGGTCA	AACAGCATT	TGCGCAAAAC	CGCGCGCCCG	AGCTTTTGGGA
901	AGCGTTTGTG	GAAAGCGTGC	GCTTTTGGG	CGAGCGCGAA	CAGCAGAAAG
951	CCATCGATT	TGCGATGCT	TGGCTGAAAG	AACAGCGCGA	TAAACGCGCT
1001	CTGCTGATGT	ATCTCGTGTG	GCTCGGCTTC	GGCGCGCAAC	TTTGGGCGAA
1051	GGCAAAAGGC	TACCTTGAAG	CGAGCATGCG	ATTAAGCGCG	AGTATTTCGC
1101	CGCGTTTGTG	TCTAAACAAAG	GTTTTCGAGC	AAATCGGAGA	ACCGCAGAAAG
1151	GGGAGGCGC	AC...			

This corresponds to the amino acid sequence <SEQ ID 750; ORF100>:

1	MKTVMVIVVL	FAAAVGLALA	SGIYTGdVVI	VLGQTLMLRN	LHAFVLGSLI
51	AVVVWYELFK	FIIGVLNIPE	KMORTGSARK	GKXKXIALNK	AGLAYEGGRF
101	EKAELEASRV	IVNVRGRNRR	TLALMLXAHX	AGOMENIXXR	DRYLAEIATKL
151	PEKQQLSRVL	LLAASALNRR	DYEAEEANHL	AAAKMANLTL	RLVRLXIRYA
201	FDRGDALQVL	AKTEKLSKAG	ALGKSEMERY	QNWAYRRQLA	DAADAALAKT
251	CLKRIPDSLK	NGELSVSVAE	KYERLGLYAD	AVKWKQHYFP	XNRRPELLEA
301	FVESVRLFGE	REQQKAIDFA	DWLKKEOPDN	ALLMLYLGRF	AFGRKLWGKA
351	KYLEASIAL	KPISARILVL	TKVFDEIGEP	QKAEAH...	

Further work revealed the complete nucleotide sequence <SEQ ID 751>:

1	ATGAAAACGG	TAGTCTGGAT	TGTCGCTCTG	TTTGCCGCGC	CGCTCGGACT
51	GGCGCTGGCT	TCGGGCATT	ACACCGGCGA	CGTGATATAC	GTACTCGGAC
101	AGACCATGCT	CAGAATCAAC	CTGCACGCT	TTGTGTTAGG	TTGCTGATT
151	GGCGCTGGCT	TGTGGTATT	CTTGTTTAAA	TTTATTATCG	GgGgTACTCA

201 TATCCCCGAA AAGATGCAGC GTTTCGGTTC GSCGCGTAAA GCGCGCAAGG
 251 CCGCGCTTGC CTTGAACAAG GCGGGTTTGG CGTATTTTGA AGGCGCTTTT
 301 GAAAGAGCGG AACTAGAAGC CTCACGCGTG TTGGTCAACA AAGAGGCGCG
 351 AGACAAACCG ACTTTGGCAT TGATGCTGGG CGCGCACGCC GCCGACACGA
 401 TGGAAACATC CGAGCTGCGC GACCGTTATC TTGCGGAAAT CCGCAACTGT
 451 CCGGAAACAT AGCAGCTTTC CGGTATCTTT TTGTTGCGCG AATCGCGCTT
 501 GAACCGCGCG GATTACGAGG CGCGGAGACG CATCTTCATC GCGCGCGCGA
 551 AGATGAATGC CAACCTTACG CGCCTCGTGC GTCTGCACT TCCTTACGCT
 601 TTGACACGGG GCGACCGGTT GCAGGTCTCT GCAAAAACCG AAAAATCTTC
 651 CAAGGCGGGC GCGTTGGGCA AATCGGAAAT GGAACGGTAT CAAAATTTGG
 701 CATACCGCGC CCAGCTGGCG GATGTCGCCG ATGCCGCGCG TTTGAAACCC
 751 TGCGTTGAAG CGATTCCCGA CAGCCTCAAA AACGGGGAAT TGAGCGTATC
 801 GGTTCGCGAA AAGTACGAAC GTTTGGGACT GTATGCCAT GCGGTCAAT
 851 GGGTCAAAAC GCATTATCCG CACAACGCCG GCCCGAGCT TTTGGAAGCC
 901 TTTGTCGAAA GCGTGCCTTT TTTGGCGGAG CGCGAACGCG AGAAAGCCAT
 951 CGATTTTGCC GATGCTTGGC TGAAGAACA GCGCGATAC GCGCTTCTGC
 1001 TGATGTATCT CGGTGCGCTC GCCTACGCCG CAAACTTTTG GGGCAAGGCA
 1051 AAAGGCTACC TTGAAGCGAG CATTGCATTA AAGCGAGTA TTTCCGCGCG
 1101 TTTGTTCTTA CAAAGGTTT TCGACGAAT CCGAGAACCG CAGAAAGCGG
 1151 AGCGCGAGCG CAATCTGGTT TTGGAAGCGG TCTCCGATGA CGAACGTCAC
 1201 GCAGCGTTAG AGCAGCATAG CTGA

This corresponds to the amino acid sequence <SEQ ID 752; ORF100-1>:

1 MKTVVIVVLFAAAVGLALASGIYTGdVYI VLQTMRLIN LHAFLVGLSLI
 51 AVVWVYFLFK FIIGVLNIP E KMORFGSARK GRKAALALNK AGLAYEGRF
 101 EKAELASRV LVNKEAGDNR TLALMLGAHA AGCMENIELR DRYLAELIAKL
 151 FEKQQLSRV LLAESALNRR DYEAEEANLH AAKMNNANIL RLVRQLRLYA
 201 FDRGDALQVL AKTEKLSKAG ALGKSEMERY QNWAYRRQLA DAADAAALKT
 251 CLKRIPDSLK NGELSVSAE KYERLGLYAD AVKWKQHYH HNRPELLEA
 301 FVESVRLFGE REQQAIDFA DAWLKEQPDN ALLIMYLRGL AYGRKLWGKA
 351 KYLEASIAL KPSISARLVL AKVFEIDGE QKAEARQNLV LEAVSDDERH
 401 AALEQHS*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF100 shows 93.5% identity over a 386aa overlap with an ORF (ORF100a) from strain A of *N.*

meningitidis:

		10	20	30	40	50	60
orf100.pep		MKTVVIVVLFAAAVGLALASGIYTGdVYI	VLQTMRLINLHAFLVGLSLI	IAVVVWYFLFK			
40	orf100a	MKTVVIVVLFAAAVGLALASGIYTGdVYI	VLQTMRLINLHAFLVGLSLI	IAVVVWYFLFK			
		10	20	30	40	50	60
		70	80	90	100	110	120
orf100.pep		FIIGVLNIP E KMORFGSARKGRKAALALNKAGLAYEGRFEKAELASRVLVNKEAGDNR					
45	orf100a	FIIGVLNIP E KMORFGSARKGRKAALALNKAGLAYEGRFEKAELASRVLVNKEAGDNR					
		70	80	90	100	110	120
		130	140	150	160	170	180
orf100.pep		TLAIMLXAHAAAGCMENIELRDRYLAEIAKLPEKQQLSRVLLAESALNRRDYEAEEANLH					
50	orf100a	TLAIMLXAHAAAGCMENIELRDRYLAEIAKLPEKQQLSRVLLAESALNRRDYEAEEANLH					
		130	140	150	160	170	180
		190	200	210	220	230	240
orf100.pep		AAAKMNNANILRLVRLXIRYAFDRGDALQVLAKTEKLSKAGALGKSEMERYQNWAYRRQLA					
orf100a		AAAKMNNANILRLVRLXIRYAFDRGDALQVLAKTEKLSKAGALGKSEMERYQNWAYRRQLX					
		190	200	210	220	230	240
		250	260	270	280	290	300
orf100.pep		DAADAAALKTCLKRIPDSLKNGELSVSAEKYERLGLYADAVKWKQHYHNRPELLEA					
orf100a		DAADAAALKTCLKRIPDSLKNGELSVSAEKYERLGLYADAVKWKQHYHNRPELLEA					
		250	260	270	280	290	300

-421-

		310	320	330	340	350	360
	orf100.pep	FVESVRFLEGEREQKAI	DFADAWLKEQPDNAL	LLMYLGRLA	FAFGRKLWGKAGKYLEASIAL		
5	orf100a	FVESVRFLEGERDQKAI	DFADAWLKEQPDNAL	LLXYLGRLAY	GRKLWGKAGKYLEASIAL		
		310	320	330	340	350	360
		370	380				
	orf100.pep	KPISISARLVLT	KVDFEIGEPQKAEAH				
10	orf100a	KPISISARLVLA	KVDFETGEPQKAEARN	LVLASVAEENR	PSAEATHX		
		370	380	390	400		

The complete length ORF100a nucleotide sequence <SEQ ID 753> is:

1	ATGAAAACGG	TAGTCTGGAT	TGTCGTCTCTG	TTTGGCGCG	CNNTCGGGCT
51	GGCATTGGCG	TCGGGCATTN	ACACCGCGCA	CGTGTATATC	GTACTCGGAC
101	AGACCATGCT	CAGAAATCAAC	CTGCAAGCCT	TTGTGTTAGG	TTGCTGTATT
151	CGCGTCGTGG	TGTGGTATTT	CTGTTCATAA	TTTCATCATG	GGGTACTCAA
201	TANCCCGGAA	AAGATGCGAG	GTTTCGGTTC	GGCGGTATAA	GGCGCAAGG
251	CCGCGCTTGC	TTTGACAAAG	GCGGGTTGG	CGTATTTTGA	AGGCGTTTT
301	GAAGAAGCGG	AACTTGAAGC	CTCGCGCTGA	TTGGGAACA	AGAAGGCGGG
351	GGATAACCGG	ACTTTGGCAT	TGATGTTGGG	CGCACATGCC	GCGGCGCAGA
401	TGGAAAACAT	CGAGCTGCGC	GACCGTTATC	TTGCGGAAAT	CGCCMAACTG
451	CCGGAAGAAC	AGCAGCTTTC	CGGTATATCT	TTGTTGGGCG	AATCGGCGTT
501	GAACCGCGCG	GATTACGAAG	CGGCGGAAGC	CAATCTTCAT	GCGGCGGCGA
551	AGATGAATGC	CAACCTTAAG	CGCCTCGTGC	GTCGCAACT	TGTTACGCT
601	TTGCGACAGG	GCGACGGT	GCAGGTTCTG	GCAAAAACG	AAAAATTTC
651	CAAGCGGGCG	GCGTNGGCA	AATCGGAAAT	GGAAACGTAT	CAAAATTGGG
701	CATACCGCGC	CCAGCTGNCG	GATGCTGCGC	ATGCGCGCGC	TTTGAAACCC
751	TGCTGAAGC	GGATTCCOGA	CAGCCTCAAA	AACGGGGAAT	TGAGCGTATC
801	GSTTGCGGAA	AAGTAGCGAAC	GTTTGGGACT	GTATCGGAT	CGGTCAGAAAT
851	GGGTCAACA	GCATTATCCG	CACACCGCC	GACCGGAAT	TTTGGAAACN
901	TTTGTGAAAA	CGCGGCTT	TTTGGCGCA	CGCATCGAG	AGAAACCAT
951	CGATTTTGGC	GATGCTTGGC	TGAAGAACA	GCGCGTAAT	CGGCTCTGTC
1001	TGANGTATCT	CGGTGGGCTC	GCTACGCGCC	GCAAACTTTG	GGGCAAGGCA
1051	AAAGGCTACC	TTGAAGCGAG	CATTGCATTA	AAGCGAGTAT	TTTCGCGCG
1101	TTTGGTCTG	GCAAGGTTT	TTGAAGCAAC	CGGAGAACCG	CAGAAGCGCG
1151	AGGCGCAGCG	CAACTTGGTT	TTGGCAAGCG	TTGCGAGGAG	AAACCGNCTT
1201	TCGCGCGAAA	CCCATTTGA			

This encodes a protein having amino acid sequence <SEQ ID 754>:

1	MKTVMVIVL	FAAAXGLALA	SGIXTGDVYI	VLQGTMLRN	LHAFVLGSLI
51	AVVWVYFLFK	FIIGVLNXP	KMQRFGSARK	GRKAALALNK	AGLAYFEGRF
101	EKAELEASRV	LGNKEAGDNR	LALMLGAHA	AGMENIELR	DRYLAEIAKL
151	FEKQQLSRYL	LLAEASALNRR	DYEAEEANLH	AAARMNANLT	RLVRLQLRYA
201	FDGDAIQVL	AKTEKSRAG	AXGSEMERI	QNWATRQLX	DAADALNLT
251	CLKRPDSLK	NSELSVSAE	KYERLGLYAD	AVKWKQHYF	HNRRPELEA
301	FVESVRFLEGE	RQCKALDEA	DAWLKEQDN	ALLXYLGR	LAFGRKLWGKA
351	KVYLEASIAL	KPISISARLV	AKVDFETGEP	QKAEARNLV	LASVAEENRP
401	SAETH*				

ORF100a and ORF100-1 show 95.1% identity in 406 aa overlap:

50	orf100a.pep	10	20	30	40	50	60
	orf100-1	10	20	30	40	50	60
55	orf100a.pep	70	80	90	100	110	120
	orf100-1	70	80	90	100	110	120
60	orf100a.pep	130	140	150	160	170	180
	orf100-1	130	140	150	160	170	180

-422-

		130	140	150	160	170	180
		190	200	210	220	230	240
5	orf100a.pep	AAAKMNAITLIRLQRYAFDRGDAIQVLAKTEKSKAGXGKSEMERYQNWAYRRQLX					
	orf100-1	AAAKMNAITLIRLQRYAFDRGDAIQVLAKTEKSKAGALGKSEMERYQNWAYRRQLA					
		190	200	210	220	230	240
10	orf100a.pep	DAADAAALKTCLKRIPOSLKNGELSVSVAEKYERLGLYADAVKWKQHYPHNRRPELLEA					
	orf100-1	DAADAAALKTCLKRIPOSLKNGELSVSVAEKYERLGLYADAVKWKQHYPHNRRPELLEA					
		250	260	270	280	290	300
15	orf100a.pep	FVESVRFGLGERDQKKAIDFADAWLKEQPDNALLXLYLGRLAYGRKLWGKARGYLEASIAL					
	orf100-1	FVESVRFGLGEREQKKAIDFADAWLKEQPDNALLMYLGRLAYGRKLWGKARGYLEASIAL					
		310	320	330	340	350	360
20	orf100a.pep	KPSISARILVLAKVFDEIGEPQKAEARNLVLASVAEENRPSA-ETHX					
	orf100-1	KPSISARILVLAKVFDEIGEPQKAEARNLVLEAVSDDERHAALQHSX					
25		370	380	390	400		
		370	380	390	400		

Homology with a predicted ORF from *N.gonorrhoeae*

ORF100 shows 93.3% identity over a 386 aa overlap with a predicted ORF (ORF100ng) from *N.gonorrhoeae*:

30	orf100.pep	MKTVMVIVVLFAAAVGLALASGIYTGDIYIVLQGTMLRINLHAFVLGSLIAVVMWYFLFK	60
	orf100ng	MKTVMVIVVLFAAAVGLALASGIYTGDIYIVLQGTMLRINLHAFVLGSLIAVVMWYFLFK	60
35	orf100.pep	FIIGVLINIPKMQRFSGARKGKXKXLAINKAGLAYFEGRFEKAELEASRVLNKVGDRNR	120
	orf100ng	FIIGVLINIPENMRRSGSARKGRKAAALINKAGLAYFEGRFEKAELEASRVLNKGAEDNR	120
	orf100.pep	TLALMLXAAHAQGMENIXRDRYLAEIAKLPEKQQLSRVLLAESALNRDRYEAEEANLH	180
40	orf100ng	TLALMLGAHAQGMENIELDRYLAEIAKLPEKQQLSRVLLAESALNRDRYEAEEANLH	180
	orf100.pep	AAAKMNAITLIRLQRYAFDRGDAIQVLAKTEKSKAGALGKSEMERYQNWAYRRQLA	240
	orf100ng	AAAKMNAITLIRLQRYAFDRGDAIQVLAKTEKSKAGALGKSEMERYQNWAYRRQMA	240
45	orf100.pep	DAADAAALKTCLKRIPOSLKNGELSVSVAEKYERLGLYADAVKWKQHYPHNRRPELLEA	300
	orf100ng	DAADAAALKTCLKRIPOSLKNGELSVSVAEKYERLGLYADAVKWKQHYPHNRRPELLEA	300
50	orf100.pep	FVESVRFGLGEREQKKAIDFADAWLKEQPDNALLMYLGRLAYGRKLWGKARGYLEASIAL	360
	orf100ng	FVESVRFGLGEREQKKAIDFADAWLKEQPDNALLMYLGRLAYGRKLWGKARGYLEASIAL	360
55	orf100.pep	KPSISARILVLKVFDEIGEPQKAEAH	386
	orf100ng	KPSIPARILVLAKVFDETAQSKAEARNLVLASVAGENRPSAETH	405

The complete length ORF100ng nucleotide sequence <SEQ ID 755> is:

1	ATGAAAACGG	TAGTCTGGAT	TGTTGCTCGT	TTTGCCGCGC	CGGTCGGACT
51	GGCGCTGGCT	TCGGGCAATT	ACACCGCGCA	CGTGTATATC	GTACTCGGAC
60	AGACCATGCT	CAGAAATCAAC	CTGCACGGCT	TTGTGTATTG	TTGCTGATT
101	CGCGTCGTGG	TGTGGTATT	CGTGTATTA	TTTCATCATCG	GGCTACTCAA
151	TATCCCGGAA	AATATGCGGC	GTTCGGGTTT	GGCGCGGAAA	GGCGCGAAGG
201	CCGCGCTTGC	CTTGAATAAG	GCGGGTTTGG	CGTATTTCGA	AGGCGGTTTT
251	GAAAGGCGGC	AAGTCGGAAG	CTCTCGAGTG	TTGGGCAACA	AAGAGGCGCG
301	AGACCAACGG	ACTTGTGGCAT	TGATGCTGGG	CGCGACGCGC	GCAGGACAGA
351	TGGAAATAT	CGAGCTGGCG	GACCTTATTC	TTGCGGAAAT	CGCCAACTGT
401					

5
10
15

```

451 CCGGAAAAAC AGCAGCTTTC CCGCTATCTT CTGCTGCGGG AATCGCGGTT
501 AAACCGGCGC GATTACGAAG CGGCGGAAGC CAATCTTCAT GCGGCGGGCA
551 AGATGAATGC CAACCTTACG CGCCTCGTGC GTCTGCAACT TCGTTACGCC
601 TTGATCGGGG GCGATGCGTT GCAGGTTCTG GCAAAAacCG AAAAATCTTC
651 CAAGCGGGCG GCGTTGGGCA AATCGGAAAT GGAACGGTAT CAAATTTGGG
701 CATACGCGCG CGAGTGGGCG GATGCTGCGG ATGCGCGCGC TTGGAAGAAC
751 TGCTTGAGAC GATTCCCGCA CAGCTCTCAA AACCGGGAAT TGagcGTATC
801 GGTTCGCGAA AAGTACGAAC GTTTGGGACT GTATGCCGAT CGGCTCAAT
851 GGGTCAAACA GCATTATCCG CACAACCGCC GCCCGGAGCT TTTGAAGGCC
901 TTTGTCGAAA GCGTGGCGCT TTTGGGCGAG CGCGAAGCAG AGAAGGCCAT
951 CGATTTTGCC GATTCTTGGC TGAAGGAACA GCCCGATAAC GCGCTTCTGC
1001 TGATGTATCT CGGCGGCGCT GCCTACGGCC GCAAACTTTC GGGTAAGGCA
1051 AAAGGCTACC TTGAAGCGAG TATTGCACTG AAGCCGAGTA TTCGCGCGCG
1101 TTTGGTGTTC GCAAAGGTTT TTGACGAAAC CGCAGAGTCG CAAAAAGCGC
1151 AAGCACAGCG CAACTTGGTT TTGGCAAGCG TTGCGCGGGA AAACCGCCCT
1201 TCCGCCGAAA CCGGTTGA

```

This encodes a protein having amino acid sequence <SEQ ID 756>:

20
25

```

1 MKTVVMIVVL FFAAVGLALA SGIYTGdVYI VLGQTMRLIN LHAFLVLSLI
51 AVVVWYLEK FIIGVLNIFP NMRSGGARK GRKAALNLA AGLAYEGRI
101 EKAELEASRV LKNKEAGDNR TLALMLGAHA AGOMENIELR DRYLAETAKL
151 PEKQQLSRVL LLAESALNRR DYEAAENLH AAKMNNANLT RLVLQLRYA
201 FDRGDALQVL AKTEKLSKAG ALGKSEMERY QNWAYRRQMA DAADPAALTK
251 CLKRIPDSLK NGELSVSAEY KYERLGLYAD AVKWKQHYFP HNNRPFELLEA
301 FVESVRFELGE REQQKAIDFA DSWLKEQPDN ALLMYLGLRL AYGRKLWGKA
351 KGYLEASIAL KPSIPARLVL AKVFDETAQS QKAEAQRLNV LASVAGENRFP
401 SAETR*

```

ORF100ng and ORF100-1 show 95.3% identity in 402 aa overlap:

30
35
40
45
50
55
60
65

```

          10      20      30      40      50      60
orf100-1.pep MKTVVMIVVLFAAAVGLALASGIYTGdVYIVLGQTMRLINLHAFLVLSLIAVVWVWYLFK
          |||
orf100ng      MKTVVMIVVLFAAAVGLALASGIYTGdVYIVLGQTMRLINLHAFLVLSLIAVVWVWYLFK
          10      20      30      40      50      60

          70      80      90      100     110     120
orf100-1.pep FIIGVLNIPKMRQFGSARKGRKAALNKAAGLAYEGRFEKAELEASRVLVNKEAGDNR
          |||
orf100ng      FIIGVLNIPENMRSGSARKGRKAALNKAAGLAYEGRFEKAELEASRVLVNKEAGDNR
          70      80      90      100     110     120

          130     140     150     160     170     180
orf100-1.pep TLALMLGAHAAGOMENIELRDRLAEIAKLPEKQQLSRVYLLAESALNRRDYEAENLH
          |||
orf100ng      TLALMLGAHAAGOMENIELRDRLAEIAKLPEKQQLSRVYLLAESALNRRDYEAENLH
          130     140     150     160     170     180

          190     200     210     220     230     240
orf100-1.pep AAAXMNNANLTRLVRLQLRYAFDRGDALQVLAKTEKLSKAGALGKSEMERYQNWAYRRQLA
          |||
orf100ng      AAAXMNNANLTRLVRLQLRYAFDRGDALQVLAKTEKLSKAGALGKSEMERYQNWAYRRQMA
          190     200     210     220     230     240

          250     260     270     280     290     300
orf100-1.pep DAADAAALKTCLKRIPDSLKNGLSVSAEYKYERLGLYADAVKWKQHYFPNNRPFELLEA
          |||
orf100ng      DAADAAALKTCLKRIPDSLKNGLSVSAEYKYERLGLYADAVKWKQHYFPNNRPFELLEA
          250     260     270     280     290     300

          310     320     330     340     350     360
orf100-1.pep FVESVRFELGEREQQAIDFADAWLKEQPDNALLMYLGLRAYGRKLWGKAGKGYLEASIAL
          |||
orf100ng      FVESVRFELGEREQQAIDFADSWLKEQPDNALLMYLGLRAYGRKLWGKAGKGYLEASIAL
          310     320     330     340     350     360

          370     380     390     400
orf100-1.pep KPSISARLVLAKEVDEIGEPOKAEAQRLNVLEAVSDDERHAALQHSX
          |||
orf100ng      KPSIPARLVLAKEVDETAQSQKAEAQRLNVLASVAGENRFPSEATRX

```

370 380 390 400

Based on this analysis, including the presence of a putative leader sequence, a putative transmembrane domain, and a RGD motif, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 90

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID 757>

```

10 1 ATGATGTTTT CTGTTTCAA GCTGTTTAC TTGTTTTTG TCATTTCGTG
51 GTTTCAGGG CTGTTTACC TGCAGAGAT TTTCGTCAAT ATGGCGATGA
101 TTGATGTGCC GCGCGGCAAT CCGAGATATG TGCGTCTGTC GGGCATGGCG
151 GTGCGGCTGT ACCGTTTAT GTGCGGTTG GGCTTCGCG CGGTCGTGTT
201 CCGCGCGGCG ATACCGTTTG CCGCGGCTG GTGGGCGAG CGCTGGGTAC
251 ACGTCAAACT GTGTTTGGG TTGATGCTCT TGCGTTACCA GTTGATTGCG
15 301 GCGGTGCTGC TGCGCCGTTT TCAGGATTAC AGCAATGCTT TTTCACACCG
351 CTGGTACCGC GTGTTCACAG AATCCCGGT GCTGCTGATG GTTGCCGCGC
401 TGTATCTGGT CGTGTTCAAA CCGTTTGA

```

This corresponds to the amino acid sequence <SEQ ID 758; ORF102>:

```

20 1 MMFSWFKLFH LFFVISWAG LFYLPRIFFN MAMIDVPRGN PEYVRLSGMA
51 VRLYRMSFL GFGAVVEGAA IPFAAGWWS GWVHVKLCLG LMLLAYQLYC
101 GVLLRRFQDY SNAFSHRWYR VFNEIPVLLM VAALYLVVEK FF*

```

Further work revealed the complete nucleotide sequence <SEQ ID 759>:

```

25 1 ATGATGTTTT CTGTTTCAA GCTGTTTAC TTGTTTTTG TCATTTCGTG
51 GTTTCAGGG CTGTTTACC TGCAGAGAT TTTCGTCAAT ATGGCGATGA
101 TTGATGTGCC GCGCGGCAAT CCGAGATATG TGCGTCTGTC GGGCATGGCG
151 GTGCGGCTGT ACCGTTTAT GTGCGGTTG GGCTTCGCG CGGTCGTGTT
201 CCGCGCGGCG ATACCGTTTG CCGCGGCTG GTGGGCGAG CGCTGGGTAC
251 ACGTCAAACT GTGTTTGGG TTGATGCTCT TGCGTTACCA GTTGATTGCG
30 301 GCGGTGCTGC TGCGCCGTTT TCAGGATTAC AGCAATGCTT TTTCACACCG
351 CTGGTACCGC GTGTTCACAG AATCCCGGT GCTGCTGATG GTTGCCGCGC
401 TGTATCTGGT CGTGTTCAAA CCGTTTGA

```

This corresponds to the amino acid sequence <SEQ ID 760; ORF102-1>:

```

35 1 MMFSWFKLFH LFFVISWAG LFYLPRIFFN MAMIDVPRGN PEYVRLSGMA
51 VRLYRMSFL GFGAVVEGAA IPFAAGWWS GWVHVKLCLG LMLLAYQLYC
101 GVLLRRFQDY SNAFSHRWYR VFNEIPVLLM VAALYLVVEK FF*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with HP1484 hypothetical integral membrane protein of *H. pylori* (accession number AE000647)

ORF102 and HP1484 show 33% aa identity in 143aa overlap:

```

40 orf102 3 FSWFKLHFLFFVISWAGLFYLPRIFFNMAMIDVPRGNPEYVRLSGMAVRLYRMSFLGF 62
FW K FH+ VISW A LFYLP R+ FV A + V++ +LY F++
HP1484 8 FLWVKAFHYIAVISWMAALFYLRFLVYHAENAHKKEFVGVIQI QEK--KLYSFIASPM 65

orf102 63 GAVVEGAAIPFAAG---WWSGSGWVHVKLCLGLMLLAYQLYCGVLLRRFQDYSNAFSHRWY 119
G + + + + GW+H KL L ++LLAY YC +R + + +R+Y
HP1484 66 GFTLLITGIIMLLIEPTLFKSGGWLHAKLALVLLLAYHFYCKCKMRELEKDPTRRNARFY 125

orf102 120 RVNEIFXXXXXXXXXXXXKPF 142
RVNE P KPF
HP1484 126 RVNEAPTILMILIVILVVVKPF 148

```

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF102 shows 99.3% identity over a 142aa overlap with an ORF (ORF102a) from strain A of *N.meningitidis*:

5	orf102.pep	MMFSWFKLH LFFVISWFAGLFYLPRI FVN MAMIDVPRGNPEYVRLSGMAVRLYRFMSPL	10	20	30	40	50	60
	orf102a	MMFSWFKLH LFFVISWFAGLFYLPRI FVN MAMIDVPRGNPEYVRLSGMAVRLYRFMSPL	10	20	30	40	50	60
10	orf102.pep	GFGAVVFGAAIPFAAGWWSGWVHV KLCGLGIMLLAYQLYCGVLLRRFDQY SNAFSHRWYR	70	80	90	100	110	120
	orf102a	GFGAVVFGAAIPFAAGWWSGWVHV KLCGLGIMLLAYQLYCGVLLRRFDQY SNAFSHRWYR	70	80	90	100	110	120
15	orf102.pep	VFNEIPVLLMVAALYL VVFKPF	130	140				
	orf102a	VFNEIPVLLMVAALYL VVFKPF	130	140				

The complete length ORF102a nucleotide sequence <SEQ ID 761> is:

1	ATGATGTTT	CTTGGTTCAA	GCTGTTTAC	TTGTTTTTG	TCATTGCTG
51	GTTTGCAGGG	CTGTTTTCAC	TGCCGAGAT	TTTCGTCAAT	ATGGCGATGA
101	TTGATGTGCC	GCGCGCAAT	CCCGAGTATG	TGCGTCTGTC	GGGCATGGCG
151	GTCGCGCTGT	ACCGTTTAT	GTCGCCGTG	GGCTTCGCG	CGGTCGTGTT
201	CGGCGCGCG	ATACCGTTT	CGCGCGCTG	GTCGCGCAG	CGCTGGGTAC
251	ACGTCAAACT	CTGTTGGCG	TTGATGCTCT	TGCGTACCA	GTGATATGC
301	GGCGTCTGC	TGCGCGCTT	TGAGGATAC	AGCAATGCT	TTTCACACG
351	CTGATACCG	GTGTTCAA	AAATCCCGT	GCTGCTGAT	GTTCGCCCGC
401	TGTATCTGT	CGTGTCAA	CGGTTTGA		

This encodes a protein having amino acid sequence <SEQ ID 762>:

1	MMFSWFKLH LFFVISWFAGLFYLPRI FVN MAMIDVPRGN PEYVRLSGMA
51	VRLYRFMSPL GFGAVVFGAA IPFAAGWWS GWVHV KLCGLG IMLLAYQLYC
101	GVLRRFDQY SNAFSHRWYR VFNEIPVLLM VAALYL VVFK PF*

ORF102a and ORF102-1 show complete identity in 142 aa overlap:

orf102a.pep	MMFSWFKLH LFFVISWFAGLFYLPRI FVN MAMIDVPRGNPEYVRLSGMAVRLYRFMSPL	10	20	30	40	50	60
orf102-1	MMFSWFKLH LFFVISWFAGLFYLPRI FVN MAMIDVPRGNPEYVRLSGMAVRLYRFMSPL	10	20	30	40	50	60
orf102a.pep	GFGAVVFGAAIPFAAGWWSGWVHV KLCGLGIMLLAYQLYCGVLLRRFDQY SNAFSHRWYR	70	80	90	100	110	120
orf102-1	GFGAVVFGAAIPFAAGWWSGWVHV KLCGLGIMLLAYQLYCGVLLRRFDQY SNAFSHRWYR	70	80	90	100	110	120
orf102a.pep	VFNEIPVLLMVAALYL VVFKPF	130	140				
orf102-1	VFNEIPVLLMVAALYL VVFKPF	130	140				

Homology with a predicted ORF from *N.gonorrhoeae*

ORF102 shows 97.9% identity over a 142 aa overlap with a predicted ORF (ORF102ng) from *N.gonorrhoeae*:

	orf102.pep	MMFSWFKLFHLEFFVISWFAGLFYLPRI FVNAMIDVPRGNPEYVRLSGMAVRLYRMSPL	60
	orf102ng	MMFSWFKLFHLEFFVISWFAGLFYLPRI FVNAMIDAPRGNPEYVRLSGMAVRLYRMSPL	60
5	orf102.pep	GFGAVVFGAAIPFAAGRWGSGVHVHVKLCGLMMLLAYQLYCGVLLRRFPQDYSNAFSHRWYR	120
	orf102ng	GFGAVVFGAAIPFAAGRWGSGVHVHVKLCGLMMLLAYQLYCGVLLRRFPQDYSNAFSHRWYR	120
10	orf102.pep	VFNEIPVLLMVAALYLVVFKPF	142
	orf102ng	VFNEIPVLLMVAALYLVVFKPF	142

The complete length ORF102ng nucleotide sequence <SEQ ID 763> is:

	1	ATGATGTTTT	CTTGTTTCAA	GCTGTTTCAAC	TTGTTTTTTG	TCATTTCTGT
15	51	GTTCGAGGG	CTGTTTTACC	TGCGAGGAT	TTTCGTCAAT	ATGGCGATGA
	101	TTGATCGCC	GCGCGCAAT	CCGAGTATG	TGCGCTGTC	GGGGATGGCG
	151	GTGCGGTGT	ACCGTTTTAT	GTCGCTTTG	GGTTTCGGCG	CGGCTGTTT
	201	CGGCGGGCG	ATACCGTTTG	CGCGggcgc	GTCGGGcagc	ggctggGTC
	251	ACGTCAACT	GTTTGCGG	TTGATGCTCT	TGCTTATCA	GTTGATATGC
	301	CGCTGTC	TGCGCTTT	TCAGGATTC	AGCAATGCTT	TTTCACACG
20	351	CTGTACTGC	GTCGTCACg	AAATCCCGCT	GCTGCTGATG	GTGCGCGCG
	401	TGTATCTGT	CGTGTCAA	CCGTTTTGA		

This encodes a protein having amino acid sequence <SEQ ID 764>:

	1	MMFSWFKLFH	LEFFVISWFA	GLFYLPRI FVN	MAMIDAPRGN	PEYVRLSGMA
25	51	VRLYRMSPL	GFGAVVFGAA	IPFAAGRWGS	GVHVHVKLCGL	MMLLAYQLYC
	101	GVLRRFPQDY	SNAFSHRWYR	VFNEIPVLLM	VAALYLVVFK	PF*

ORF102ng and ORF102-1 show 98.6% identity in 142 aa overlap:

		10	20	30	40	50	60
	orf102-1.pep	MMFSWFKLFHLEFFVISWFAGLFYLPRI FVNAMIDVPRGNPEYVRLSGMAVRLYRMSPL					
30	orf102ng	MMFSWFKLFHLEFFVISWFAGLFYLPRI FVNAMIDAPRGNPEYVRLSGMAVRLYRMSPL					
		10	20	30	40	50	60
	orf102-1.pep	GFGAVVFGAAIPFAAGRWGSGVHVHVKLCGLMMLLAYQLYCGVLLRRFPQDYSNAFSHRWYR					
35	orf102ng	GFGAVVFGAAIPFAAGRWGSGVHVHVKLCGLMMLLAYQLYCGVLLRRFPQDYSNAFSHRWYR					
		70	80	90	100	110	120
	orf102-1.pep	VFNEIPVLLMVAALYLVVFKPEX					
40	orf102ng	VFNEIPVLLMVAALYLVVFKPEX					
		130	140				
	orf102-1.pep	VFNEIPVLLMVAALYLVVFKPEX					
	orf102ng	VFNEIPVLLMVAALYLVVFKPEX					
		130	140				

In addition, ORF102ng shows significant homology to a membrane protein from *H. pylori*:

45	gi 2314656 (AE000647) conserved hypothetical integral membrane protein		
	[Helicobacter pylori] Length = 148		
	Score = 79.2 bits (192), Expect = 1e-14		
	Identities = 50/147 (34%), Positives = 68/147 (46%), Gaps = 13/147 (8%)		
50	Query: 3	FSWFKLFHLEFFVISWFAGLFYLPRI FVNAMIDAPRGNPEYVRLSGMAVRLYRMSPLGF	62
	Sbjct: 8	FLWVKAFHVIIVISWMAALFYLPRI FVYHAENAHKKEFVGQVQIQEK---KLVSFIASPM	65
		G + + + F + G GW+H KL L ++LLAY YC +R + +	
55	Query: 63	GAVVFGAAIP-----FAAGRWGSGVHVHVKLCGLMMLLAYQLYCGVLLRRFPQDYSNAFS	115
	Sbjct: 66	GFTLTITGIMLLIEPTLFKSG---GWLHAKLALVLLLAYHFYCKKCMRELEKDPTRN	121
		R+YRVNE P KFF	
60	Query: 116	HRVYRVFNEIPXXXXXXXKXKFF	142
	Sbjct: 122	ARFYRVFNEAFTILMILIVLVVVKFF	148

Based on this analysis, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 91

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 765>:

```

5      1 ATGCGAAAAA TGATGAAATG GCGCGGTGTT GCGCGCGTCG CGGCGGCAGC
51     GGT'TTGGGGC GGGTGGTCTT AACTGAAGCC CGAGCGCGAC GTGCTTGATA
101    TTAACGGAAC GGTCAAGCGC GGC // .....
//... ATTTGCTTTA CGATTTTGTG CGAACGGAT ACGCCGATTA AGCGGAAGCT
51     CGACAGCGTC GACCCCGGGC TGACACGAT GTCTGCGGCG GGTTCACACA
10     CGAGTACGGA TACGGCTTCC AATCGGCTCT ACTATTATGC CCGTTCGTTT
151    GTGCCGAATC CGACGCGCAA ACTCGCCACG GGGATGACGA CGCAAGAATC
201    GGT'TGAAATC GACGGCGTGA AAAATGTGCT GATTATTCCG TCGCTGACCG
251    TGAATAATCG CGGCGGCAAG GCGT'TTGTG GCGTGT'TGGG TCGCGGACGCG
301    AAGCGCGCGG AACGCGAAAT CGGACCGGTT ATGAGAGACA GTATGAATAC
15     CGAAGTAAAA AGCGGGT'TGA AAGAGGGGGA CAAAGTGTGC ATCTCCGAAA
401    TAACGCGCGC CGAGCAACAG GAAAGCGGCG AACGCGCCCT AGGCGGCGCG
451    CCGCGCCGAT AA

```

This corresponds to the amino acid sequence <SEQ ID 766; ORF85>:

```

20      1 MAKMMKWAIV AAVAAAAVWG GWS.LKPEPH VLDITETVRR G.....
51     .....
101    .....
151    .....
201    .....
251    PIKAKLDSVD PGLTMMSSGG YNNSDITASN AVYYIARSGV PNPDKLATG
301    MTTQNTVEID GVKMVLIIPI LTVKNRGGKA FVRVLGADGK AAEREIRTMG
351    RDSMNTVEKS GLKELGKVI SEITRAEQQE SGERALGPP RR*

```

Further work revealed the further partial nucleotide sequence <SEQ ID 767>:

```

30      1 ..GTATCGGTG GCGCGCAGGC ATOSGGGCGC ATTAAGATAC TTTATGTCAA
51     ACTCGGGCAA CAGGTTAAAA AGGGCGATT GATTGCGGAA ATCAATTGCA
101    CCTCGCAGAC CAATACGCTC AATAOGGAAA AATCCAAGTT GGAAACGTAT
151    CAGGCGAAGC TGGTGTGCGC ACAGATTGCA TTGGGCGAGC CGGAGAGAAA
201    ATATAAGCGT CAGGCGCGGT TATGGAAGGA AAGCGGACT TCCAAAGAGG
251    ATTTGGAAGG CGGCGAGGAT GGGTTTGGCG CGGCAAAGC CAATGTTGCC
301    GAGCTGAAGG CTTTAACTAG ACAGAGCAAA ATTTCCATCA ATACGCGCGA
35     GTCGGAATGG GCTTACACCG GCATTACCG ACAGATGAGC GGCACGGTGG
401    TGCGGATCTT CCGTGAAGAG GCGCGAGCT TGAGCGGCG CAGCTGTACG
451    CCGACGATTTG TCCAATTGGC GAATCTGGAT ATGATGTTGA ACAAAATGCA
501    GATTGCGCAG GCGCATATTA CCAAGTGA AAAGGGGCGC GATATTTCCG
551    TTACGATTTT GTCCGAACCG GATACGCGGA TTAAGGCGAA GCTCGACAGC
601    GTCGACCCCG GCGTGACCAC GATGTCGTG GCGGTTTACA ACAGCAGTAC
651    GGATACGGCT TCCAATGCGG TCTACTATTA TGCCGTTTGC TTTGTGCCGA
701    ATCGGACGCG CAAACTCGCC ACGGGGATGA CGAGCGAGAA TAOGGTTGAA
751    ATCGACGGCG TGAATAATGT GCTGATTATT CCGTCTCGTA CCGTGAAGAA
801    TCGCGGCGCG AAGCGGTTTG TCGCGGTGTT GGGTGGCGAG GGCAGGCGCG
45     CGGAACGCGA AATCGGACC GGTATGAGAG ACAGTATGAA TACCGAAGTA
901    AAAAGCGGGT TGAAGAGGGG GGACAAAGTG GTCATCTCCG AAATAACCGC
951    CGCGGAGCAA CAGGAAAGCG GCGAACGCGC CTAGCGCGC CCGCGCGCGC
1001   GATAA

```

This corresponds to the amino acid sequence <SEQ ID 768; ORF85-1>:

```

50      1 ..VSVGAQASQG IKILYVKLQG QVKKGLDIAE INSTSQNTLI NTEKSKLETY
51     QAKLVSAQIA LGSAEKKYKR QAALWKNENAT SKEDLESQD AFAAAKANVA
101    ELKALIRQSK ISINTAESEL GYTRITATMD GTVVAILVEE GGTVNAQST
151    PTIVOLANLD MMLNMQIAIE GDITKVKAGQ DISFTILSEP DTPIKAKLDS
201    VDFPLTMMSS GGYNSSTDTA SNAVYYIARS FVNPFGKLA TGMTTQNTVE
55     IDGVKNVLII PSLTVKNRGG KAFVRLVGLD GKAAEREIRT GMRDSMNTVE
301    KSLKELGDKV VISEITAAEQ QESGERALGG PPRR*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF85 shows 87.8% identity over a 41aa overlap and 99.3% identity over a 153aa overlap with an ORF (ORF85a) from strain A of *N. meningitidis*:

5	orf85.pep	MAKMMKWA	10	20	30	40	
	orf85a	MAKMMKWA	10	20	30	40	50 60
10	orf85.pep	80	90	100		
	orf85a	80	90	100		
15	orf85.pep	110	120	130	140	150 160
	orf85a	110	120	130	140	150 160
20	orf85.pep	170	180	190	200	210 220
	orf85a	170	180	190	200	210 220
25	orf85.pep	230	240	250	260	
	orf85a	230	240	250	260	
30	orf85.pep	270	280	290	300	310 320
	orf85a	270	280	290	300	310 320
	orf85a	330	340	350	360	370 380
	orf85a	330	340	350	360	370 380
	orf85a	390				

The complete length ORF85a nucleotide sequence <SEQ ID 769> is:

35	1	ATGGCAAAAA	TGATGAATG	GCGCGCTGTT	GCGCGCGCTG	GCGCGCGCAG
	51	GGTTTGGGGC	GGATGGTCTT	ATCTGAAGCC	CGAGCCGCGAC	CGTGTCTATA
	101	TTACGGAAAC	GGTCAGGCGC	GCGGACATCA	GCGGACAGGT	TTCTGCAACA
	151	GGGGAGATT	CGCGTCCAA	CCTGSTATCG	GTCGCGCGCG	AGGCATCGGG
	201	GCAGATTAA	AACTTTATG	TCAAACTCGG	GCAACAGGT	AAAAAGGCG
	251	ATTTGATTGC	GGAAATCAAT	TGCACTTCGC	AGACCAATAC	GCTCAATACG
40	301	GAATAATCCA	AATTGGAAAC	GTATCAGGCG	AAGCTGGTGT	CGCGCATAGT
	351	TGCTATTGGC	AGCGCGAGA	AGAATAATTA	GCGTCAGGCG	CGCTTGTGGA
	401	AGGATGATGC	GACCGCTAAA	GAAGATTTCG	AAAGCGGACA	GGATCGCTCT
	451	GCGCGGCCCA	AAGCCATGT	TGCGAGCTGC	AAGCTCTAA	TCAGACAGAG
	501	CAAAATTTC	ATCAATACG	CGAGTCGGA	ATTGGGCTAC	ACGGGCATTA
45	551	CCGCAACGAT	GGACGGCAG	GTGGTGCGA	TTCTCGTGA	AGAGGGCAG
	601	ACTGTGAAGC	CGGCGGATC	TACGCGAAG	ATTGTCCAAT	TGGCGAATCT
	651	GGATATGATG	TTGAACAAA	TGCAGATTGC	CGAGGGCGAT	ATTACCAAGG
	701	TGAAGCGGGG	CGAGGATATT	TGCTTTACGA	TTTTGTCCGA	ACCGGATACG
	751	CCGATTAAAG	CGAAGCTCGA	CAGGCTCGAC	CCGCGGCTGA	CCACGATGTC
50	801	GTCGGGCGCG	TACAACAGCA	GTACGGATAC	GGCTTCCAAT	CGGCTCTACT
	851	ATTATGCGCG	TTGCTTTGTG	CGGAATCCGG	ACGGCAAACT	CGCCACGGGG
	901	ATGACGACGC	AGAATACGCT	TGAATCGCAG	GGTGTGAAAA	ATGTGCTGAT
	951	TATTCCGCTG	CTAGCCGTGA	AAAATCGCGG	CGGCGAGGCG	TTTGTGCGCG
	1001	TTTGTGGTGC	AGACGCGAAG	CGCGCGAAG	CGGAATCCG	GACCGTATG
55	1051	AGAGACAGTA	TGAATACCGA	AGTAAAGG	GGGTTGAAG	AGGGGACAA
	1101	AGTGGTCATC	TCCGAATAAA	CCGCGCGCGA	CGACGAGGAA	AGCGCGAAG
	1151	GCGCCCTAGG	CGGCGCGCG	CGCCGATAA		

This encodes a protein having amino acid sequence <SEQ ID 770>:

60	1	MAKMMKWA	AAVAAAAVWG	WWSYLKPEPQ	AAYITETVRR	GDISRTVSAT
	51	GETSFENLVS	VGAQASGGIK	KLYVKLQQV	KKGDLIAEIN	STSQNTNLT
	101	EKSLETYQA	KLVSQIALG	SAEKYKQRA	ALWKDDATAK	EDLSAQDAL
	151	AAAKANVVEL	KALIRQKSI	TNTAESELG	TRITATMDGT	VVAIVVEESG
	201	TVNAQSTPT	IVQLANLDM	LNKMQIAEGD	ITKVKAGQDI	SFTILSEPTD
	251	PIAKLDSVD	PGLTMSMSG	YNSSTDASN	AVYYARSFV	PNPDGKLATG
65	301	MTTQNTVEID	GVMNVLIIPS	LTVMIRGGRA	FVRLVGADGK	AAEREIRTMG

351 RDSMNTVEVKS GLKEGDKVVI SEITAAEQE SGERALGGPP RR*

ORF85a and ORF85-1 show 98.2% identity in 334 aa overlap:

5	orf85a.pep	30	40	50	60	70	80
	orf85-1	PQAA	YITETVRRGDISRTVSATGEISF	SNLVS	VGAQASGQIKKLYVKLGQ	QVKKGDLIAE	
10	orf85a.pep	90	100	110	120	130	140
	orf85-1	INSTSCTNTLNTEKSKLETYQAKLVSAQIALGSAEKKYKRC	AAALWKDDATAKEDLES	QAQD			
15	orf85a.pep	150	160	170	180	190	200
	orf85-1	ALAAKANKVAELKALIRQSKISINTAESELGYTRITATMDGTVVAILVEEGQTVNAAQST					
20	orf85a.pep	210	220	230	240	250	260
	orf85-1	PTIVQLANLDMMLNKMQIAEGDITKVKAGQDISFTILSEPDTPIKAKLDSVDPGLTMS					
25	orf85a.pep	270	280	290	300	310	320
	orf85-1	GGYNSSTDTSANAVYYARSFVNPDKLATGMTTQNTVEIDGVKNVLIIPSLTIVKNRGG					
30	orf85a.pep	330	340	350	360	370	380
	orf85-1	RAFVRVLGADGKAAEREIRTMGRDSMNTVEVKSGLKEGDKVISEITAAEQESGERALGG					
35	orf85a.pep	390	400	410	420	430	440
	orf85-1	PFRRX					
40	orf85a.pep	450	460	470	480	490	500
	orf85-1	INSTSCTNTLNTEKSKLETYQAKLVSAQIALGSAEKKYKRC					
45	orf85a.pep	510	520	530	540	550	560
	orf85-1	AAALWKDDATAKEDLESQAQD					
50	orf85a.pep	570	580	590	600	610	620
	orf85-1	PTIVQLANLDMMLNKMQIAEGDITKVKAGQDISFTILSEPDTPIKAKLDSVDPGLTMS					
55	orf85a.pep	630	640	650	660	670	680
	orf85-1	GGYNSSTDTSANAVYYARSFVNPDKLATGMTTQNTVEIDGVKNVLIIPSLTIVKNRGG					
60	orf85a.pep	690	700	710	720	730	740
	orf85-1	RAFVRVLGADGKAAEREIRTMGRDSMNTVEVKSGLKEGDKVISEITAAEQESGERALGG					
65	orf85a.pep	750	760	770	780	790	800
	orf85-1	PFRRX					

Figure 19D shows plots of hydrophilicity, antigenic index, and AMPHI regions for ORF85a..

Homology with a predicted ORF from *N.gonorrhoeae*

45 ORF85 shows a high degree of identity with a predicted ORF (ORF85ng) from *N.gonorrhoeae*:

ORF85	1	MAKMMKWA	AAVAAAAVWGGWS	LKPEPHVLDITETVRRG	40
ORF85ng	1	MAKMMKWA	AAVAAAAVWGGWSY	LKPEPQAA	YITEAVRRGDISRTVSAT	50
ORF85	ISFTILSEPD	250
ORF85ng	201	TVNAAQSTPTIVQLANLDMMLNKMQIAEGDITKVKAGQDISFTILSEPD	250
ORF85	251	PIKAKLDSVDPGLTMSGGYNSSTDTSANAVYYARSFVNPDKLATG	300
ORF85ng	251	PIKAKLDSVDPGLTMSGGYNSSTDTSANAVYYARSFVNPDKLATG	300
ORF85	301	MTTQNTVEIDGVKNVLIIPSLTIVKNRGGKAFVRVLGADGKAAEREIRTM	350
ORF85ng	301	MTTQNTVEIDGVKNVLIIPSLTIVKNRGGKAFVRVLGADGKAAEREIRTM	350
ORF85	152	RDSMNTVEVKSGLKEGDKVISEITAAEQESGERALGGPFR	393
ORF85ng	351	KDSMNTVEVKSGLKEGDKVISEITAAEQESGERALGGPFR	393

The complete length ORF85ng nucleotide sequence <SEQ ID 771> is:

```

1  ATGGCGAAAAA  TGATGAAATG  GCGGGCTGTT  GCGGCGGTCG  CGGCGGCaac
51  GGTTTGGGGC  GGATGCTCTT  ATCTGAAGCC  CGAACCGCAG  GCTGCTTATA
101  TTACGGAaacc  ggTCAGGCGC  GCGGATATCA  GCGGACCGGT  TTCCGCGACG
151  GgcaGATATT  CGCCGTCGCA  CCTGGTATCG  GTCGCGCGCG  AGGCTTCGGG
201  GCAGATTAAA  AAGCTTTATG  TCAAACTCGG  GCAACAGGCG  AAAAAGGGCG
251  ATTTGATTGC  GGAATCAAT  TCGACACGCG  AGACCAACAC  GATCGATATG
301  GAAAAATCCA  AATTGGAAAC  GTATCAGGCG  AAGCTGGTGT  CGGCACAGAT
351  TGCATTGGCG  AGCGCGGAGA  AGAATATATA  CGCTCAGGCG  CGGTTTGGGA
401  AGGATGATGC  GACCTCTAAA  GAAGATTTGG  AAAGCGGCGA  GGAATGCGCTT
451  GCGCGCGGCC  AAGCCAATGT  TGCCGAGTTG  AAGGCTTTAA  TCAGACAGAG
501  CAAAATTCC  ATCAATACCG  CGAGTTCGGA  TTTGGGCTAC  ACGCGAATTA
551  CCGCGACGAT  GCGCGGCACT  FTGGTGGGGA  TTCCGCTGGA  AGAGGGCAG
601  ACTGTGAACG  CGCGCGCATC  TACGCGGAGT  ATTGTCCAAT  TGGCGATCTT
651  GGAATGATG  TTGAACAAAA  TGCAGATTGC  CGAGGCGGAT  ATTACCAAGG
701  TGAAGCGCGG  CGAGGATATT  TCGTTTACGA  TTTTGTCCGA  ACCGGATACG
751  CGGATTAAAG  CGAAGCTCGA  CAGCGTCGAC  CCGGGGCTGA  CCACGATGTC
801  GTCGCGCGCG  TACAACAGCA  GTACGGATAC  GGCTTCCAAT  CGGCTCTATT
851  ATTAATGCGG  TTGCTTTGTG  CCGAATCGGG  ACGGCAAACT  CGCCACGGGG
901  ATGACGACGC  AGAATACGGT  TGAATTCGAC  GGTGTGAAAA  ATGTGTGTCT
951  TATTCCGTCG  CTGACCGTGA  AAAATCGCGG  CGGCAAGGCG  TTCGTACGCG
1001  TGTGGGTGTC  GGACGGCAAG  GCAGTGGAAC  GCGAAATCCG  GACCGGTATG
1051  AAGAGACAGTA  TGAATACCGA  AGTGAAGAAC  GGGTTGAAG  AGGGGGACAA
1101  AGTGGTCATC  TCGGAAATAA  CGCGCGCCGA  GCAGCAGGAA  AGCGGGCAAC
1151  GCGGCTTAGG  CGGCGCGCGC  CGCCGATAA

```

This encodes a protein having amino acid sequence <SEQ ID 772>:

```

1  MAQMMKWAAR  AAVAAAAMVG  GWSYKLPEPQ  AAYITEAVRR  GDISRTVSAT
51  GEISPSNLVS  VQAQASGQIK  KLYVKLGQDV  KKGDLIAEIN  STTQNTNIDM
101  EKSCLKETYQA  KLVSAQIALG  SAEKKYKQRA  ALWKDDATSK  EDLESAQDAL
151  AAKANVAEL  KALIRQSKIS  INTAESDLGY  TRITATMDGT  VVAIPVEEGQ
201  TVNAAQSTPT  IVOLANLMM  LNKMIAGD  ITKVKAGQDI  SFTILSEPTD
251  PIKAKLDSVD  PGLTMTSSGG  VNSSTDASN  AVYYARSEFV  PNPDGKLATG
301  MTTQNTVEID  GVKNVLLIPS  LTVKNRGGKA  FVRVLGADGX  AVEREIRTMG
351  KDMNTEVKS  GLKEGDKVVI  SEITAAEQE  SGERALGGFP  RR*

```

35 ORF85ng and ORF85-1 show 96.1% identity in 334 aa overlap:

```

          30      40      50      60      70      80
orf85ng  PQAAAYITETVRGRDISRTVSATGEISPSNLVSVGAQASGQIKKLYVKLGQQVKKKGDLIAE
          |||
orf85-1  VSVGAQASGQIKLIVKLGQVKKKGDLIAE
          10      20      30

          90      100     110     120     130     140
orf85ng  INSTTQNTNTDMEKSKLETYQAKLVSAQIALGSAEKKYKQRAALWKDDATSKEDLESAQD
          |||:||||: |||
orf85-1  INSTSQNTNINTEKSKLETYQAKLVSAQIALGSAEKKYKQRAALWKENATSKEDLESAQD
          40      50      60      70      80      90

          150     160     170     180     190     200
orf85ng  ALAAAKANVAELKALIRQSKISINTAESDLGYTRITATMDGTVVAIPVEEGQTVNAAQST
          |||:||||: |||
orf85-1  AFAAAKANVAELKALIRQSKISINTAESELDGYTRITATMDGTVVAILVEEGQTVNAAQST
          100     110     120     130     140     150

          210     220     230     240     250     260
orf85ng  PTIVOLANLDMMLNKMIMIAEGDITKVKAGQDISFTILSEPTPIKAKLDSVDPLTMTSS
          |||
orf85-1  PTIVOLANLDMMLNKMIMIAEGDITKVKAGQDISFTILSEPTPIKAKLDSVDPLTMTSS
          160     170     180     190     200     210

          270     280     290     300     310     320
orf85ng  GGYNSSTDASNNAVYYARSEFVFNPDGKLATGMTTQNTVEIDGVKNVLLIPSLTVKNRGG
          |||
orf85-1  GGYNSSTDASNNAVYYARSEFVFNPDGKLATGMTTQNTVEIDGVKNVLLIPSLTVKNRGG
          220     230     240     250     260     270

          330     340     350     360     370     380

```

```

orf85ng      KAFVRLVGADGKAVEREIRTGMKDSMNTVEVKSGLKEGDKVVEISEITAAEQESGERALGG
|||||
orf85-1      KAFVRLVGADGKAAREIRTGMKDSMNTVEVKSGLKEGDKVVEISEITAAEQESGERALGG
                280      290      300      310      320      330

5
orf85ng      390
P PRRX
|||||
orf85-1      P PRRX

```

10 In addition, ORF85ng shows significant homology to an *E. coli* membrane fusion protein:

```

gi|1787104 (AEC00189) c380; 27% identical (27 gaps) to 332 residues from
membrane fusion protein precursor, MTRC_NEICO SW: P43505 (412 aa) [Escherichia
coli] Length = 380
Score = 193 bits (485), Expect = 2e-48
15 Identities = 120/345 (34%), Positives = 182/345 (51%), Gaps = 13/345 (3%)

Query: 29 PQAAVITETVRRGDISRTVSATGEISPNLVSQGAQSGQIKKLVKLGQVKKGDIAE 88
P Y T VR GD+ ++V ATG++ V VQAQ SQ+K L V +G +VRK L+
20 Sbjct: 41 PVPTFYQT LIVRPGDLQGSVLATCKLDALRKVDVGAQVSGQLKTLVAIGDKVKKDQLGV 100

Query: 89 INSTTQTNTIDMEKSKLETYQAKLVSAQIALGSAEKYKRAALWKDDATSKEXXXXXX 148
I+ N I ++ L +A+ A+ L A Y RQ L + A S++
25 Sbjct: 101 IDPEQAENQIKEVEATLMELRAQRQQAELKRLAVTYSRQRLAQTKAVSQDLDATA 160

Query: 149 XXXXXXXXXXXXXXXXIRQSKISINTAESDLGYTRITATMDCTVVAIPVEEGOTVNAQST 208
I+++ S++TA+++L YTRI A M G V I +GOTV AAO
30 Sbjct: 161 EMAYKQAQIGTIDAQIKRNQASLDATAKNTLDYTRIVAPMAGEVTQITTLQGOVIAAQA 220

Query: 209 PTIVQLANLDMMLNMQIAEGDITKVKAGQDISFTILSEPDPTEIKAKLDSVDPLGTTMES 268
P I+ LA++ ML K Q++E D+ +K GQ FT+L +P T + ++ V P
35 Sbjct: 221 PNIITLADMSAMLVKAQVSEADVTHLKPCKQAWFTVLGDLPLTRYEGQIKDLVLE----- 273

Query: 269 GCYNSSTDASNAYYYYARSFVFNPDCKLATGMTNTQNTVEIDGVKNVLLIPSLTVKNRGC 328
+ + ++A++YYAR VFN+G L MT Q ++ VKNVL IP ++ G
40 Sbjct: 274 -----TPEKYNDAIFYARFEVFNENGLRLDHTAQVHQLTQTVKNVLLIPLSALGDVPG 328

Query: 329 KAFVRLV-LGADGKAVEREIRTGMKDSMNTVEVKSGLKEGDKVVEISE 372
+V L +G+ ERE+ G ++ + E+ GL+ GD+vvi E
45 Sbjct: 329 DNRVYVKLLRNGETREREVITGARNDTDVEIVKLEAGDEVVIGE 373

```

40 Based on this analysis, it was predicted that the proteins from *N. meningitidis* and *N. gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF85-1 (40.4kDa) was cloned in the pGex vectors and expressed in *E. coli*, as described above.

The products of protein expression and purification were analyzed by SDS-PAGE. Figure 19A shows the results of affinity purification of the GST-fusion protein. Purified GST-fusion protein was used to immunise mice, whose sera were used for Western blot (Figure 19B), FACS analysis (Figure 19C), and ELISA (positive result). These experiments confirm that ORF85-1 is a surface-exposed protein, and that it is a useful immunogen.

Example 92

The following partial DNA sequence was identified in *N. meningitidis* <SEQ ID 773>:

```

50      1 ..ATTCCCGCCA CGATGACATT TGAACGCAGC GGCAATGCTT ACAAATCGT
51      TTGCGAGATT AACTGCCCG TATACAAATAT CCGTTTCGAG TCCGCGGGTA
101     CGGTTGTCGG CAATACCTCG CACCCTACCT ACTATACAGA CATACGCGAG
151     CGGCAACTCT ATGCGGAGac CAATTCCGCC GACGcGAGG TACTTACCGG
201     CAAAGCGGGC GAGAGCAAAA CCGAGCAAGC CCCCAGGGCT ATGCGTTTGT

```

-432-

251 TCACGCTTGC CTGGCAGTTG GGGCAAATG ACGGAAACT CCCCCGGGG
 301 CTGAAAATCA CCAACGGCAA AAACTTTAT TCCGTGGCG GTTTGAATAA
 351 GCGGGGTACA GGAAAATACA GCATAGCGCG CGTGAAACC GAAGTCGTCA
 401 AATATCGGGT GCGGCGCGCG GACGATCGCG TAATGATTT CTTCGCCCGG
 451 TCCTTGAACA ATATTCCGGC ACAAAATCGG TATACGCG ACGGCAAAAC
 501 CTATACGCTG AAACCTCAAT CGGTGCAGAT CACGCGCAG CGACGCAAAAC
 551 CGTAA

This corresponds to the amino acid sequence <SEQ ID 774; ORF120>:

1 ..IPATMTFERS GNAYKIVSTI KVPLYNIRFE SGGTVVGNLT HPTYYRDIRR
 51 GKLYAEAKFA DGSVTYKGAG ESKTEQSPKA MDLFTLAWQL AANDAKLPPG
 101 LKITNGKKLY SVGGINKAGT GKYSIGGVET EVVKYRVRG DDAVMYFFAP
 151 SLNNIPAQIG YTDGKTYTL KLKSVQINGQ AAKP*

Further work revealed the complete nucleotide sequence <SEQ ID 775>:

1 ATGATGAAGA CTTTAAAAA TATATTTTC GCGCCATTT TGTCGCGCGC
 51 CCGCGCGTGC GCGTATGCGG CAGGGCTGCC CCAATCGCG GTGCTGCACT
 101 ATTCGCGCAG CTACGGCATT CCGGCCAAGA TGACATTGA ACGCAGCGGC
 151 AATGCTTACA AAATGTTTT GACGATTAAA GTGCGCGTAT ACAATATCGG
 201 TTTTCGAGTCC GGGGTACGG TTGTGCGCA TACCCTGCAC CTAACCTACT
 251 ATAGAGACAT ACGCAGGGGC AAACGTGATG CGGAAGCCAA ATTGCGCCAG
 301 GCGCGCTTAA CTTACGGCAA AGCGGGCGAG AGCAAAACG AGCAAAAGCC
 351 CAGGCTATG GATTGTTCA GCTTGCCTG CAGGTGGG GCAAAAGAC
 401 CGAACTCCC CCGGGGCTG AAAATCACA ACGGCAAAA ACTTATTCG
 451 GTCGGCGGTT TGAATAAGGC GGGTACAGA AAATCAGCA TAGCGCGGT
 501 GGAACCGAA TGCTCAAAT ATCGGGTGG GCGCGCGGAC GATGCGGTAA
 551 TGTATTTCTT CGCACCGTCC CTGAACATA TTCGGGCACA AATCGGCTAT
 601 ACGGACGAGC GCAAAACCTA CTGCTGAAA CTCAAATCGG TGCAGATCAA
 651 CGGCCAGGCA GCCAAACCGT AA

This corresponds to the amino acid sequence <SEQ ID 776; ORF120-1>:

1 MMKTFKNIFS AAILSAALPC AYAAGLPQSA VLHYSYSGYI PATMTFERSG
 51 NAYKIVSTIK VPLYNIRFES GGTVVGNLT HPTYYRDIRG KLYAEAKFAD
 101 GSVTYKGAGE SKTEQSPKAM DLFTLAWQLA AANDAKLPPGL KITNGKKLYS
 151 VGGINKAGTG KYSIGGVETE VVKYRVRGD DAVMYFFAPS LNNIPAQIGY
 201 TDDGKTYTLK LKSVQINGQA AKP*

Computer analysis of this amino acid sequence gave the following results:

35 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF120 shows 92.4% identity over a 184aa overlap with an ORF (ORF120a) from strain A of *N. meningitidis*:

						10	20	30
40	orf120.pep					IPATMTFERSGNAYKIVSTIKVPLYNIRFE		
	orf120a	SAAILSAALPCAYAGLPKSAVLHYSYSGYIPATXXXXXXXXXNAYKIVSTIKVPLYNIRFE						
		10	20	30	40	50	60	
45	orf120.pep		40	50	60	70	80	90
	orf120a	SGGTVVGNLTLPHTYYRDIRRGGKLYAEAKFADGSVTYKGAGE SKTEQSPKAMD LFTLAWQL						
		70	80	90	100	110	120	
50	orf120.pep		100	110	120	130	140	150
	orf120a	AANDAKLPPGLKITNGKKLYSVGGINKAGTGKYSIGGVETE EVVKYRVRGDDAVMYFFAP						
		130	140	150	160	170	180	

55

-433-

```

              160      170      180
orf120a.pep  SLNNI PAQIGYTDGKTYTLKLKSVQINGQAAKFX
              |||
orf120a      SLNNI PAQIGYTDGKTYTLKLKSVQINGQAAKFX
              190      200      210      220

```

The complete length ORF120a nucleotide sequence <SEQ ID 777> is:

```

1   ATGATGAAGA CTTTAAATAA TATATTTC GCGGCATTT TGTCCGCCG
51  COTGCCGTGC GCGTATGCGG CAGGGCTGCG CNAATCGCGC GTGCTGCACT
101 ATTCCGCGCAG CTACGCGCATT CCGCCACANA NNANNTNNGN ACNNNGNGNC
151 AATGCTTACA AAATCGTTTC GAGCATTAJA GTGCGCGTAT ACAATATCCG
201 TTTGAGTCCG GCGGCTACGG TTGTGCGCAA TACCCTGCAC CCTACCTACT
251 ATAGAGACAT ACGCAGGGGC AAACGTGATG CGGAAGCCAA ATTGCGCGAC
301 GGCAGCGTAA CCTACGCGCA AGCGGNNNNN ANCNNNNNNG NGCAAAGCCC
351 CAAGGCTATG GATTGTGTCA CGCTTGCNTG GCGAGTGGCG GCAAAATGACG
401 CGAAATCCCG CCGGGGCGTG AAAATCACA ACGCCAAAAA ACTTTATTCC
451 GTGCGCGGTT TGAATAAGCG GGGTACAGCA AAATACAGCA TAGGCGCGGT
501 GGAACCGCAA GTGCTCAAA ATCGGCTGCG GCGCGCGGAC GATGCGGTAA
551 TGTATTTCCT CGCACCGTCC CTGAACATA TTCCGSCCA AATCGGCTAT
601 ACGACGACG GCAAAACCTA TACGCTGAAA CTCAAATCGG TGCAGATCAA
651 CGGCCAGGCA GCCAAACCGT AA

```

This encodes a protein having amino acid sequence <SEQ ID 778>:

```

1   MMKTFKNIFS AAILSAALPC AYAAGLPXSA VLHYSGSYGI PATXXXXXXX
51  NAKKIVSTIK VPLYNIRFES GGTVVGNLH PTYRDIRRG KLYAEAKFAD
101 GSVTYGKAXX XXXXQSPKAM DLFTLAWQLA ANDAKLPPGL KITNGKKLYS
151 VGLNKGATG KYSIGGVETE VVKYRVRGD DAVMYFFAPS LNNI PAQIGY
201 TDDGKTYTLK LKSVQINGQA AKF*

```

ORF120a and ORF120-1 show 93.3% identity in 223 aa overlap:

```

              10      20      30      40      50      60
orf120a.pep  MMKTFKNIFS AAILSAALPCAYAAGLPXSAVLHYSGSYGI PATXXXXXXXNAKKIVSTIK
              |||
orf120-1      MMKTFKNIFS AAILSAALPCAYAAGLPQSAVLHYSGSYGI PATMTFFERSGNAYKIVSTIK
              10      20      30      40      50      60

              70      80      90      100     110     120
orf120a.pep  VPLYNIRFESGGTVVGNLHPTYYRDIRRGKLYAEAKFADGSVITYGKAXXXXXXQSPKAM
              |||
orf120-1      VPLYNIRFESGGTVVGNLHPTYYRDIRRGKLYAEAKFADGSVITYGKAGESKTEQSPKAM
              70      80      90      100     110     120

              130     140     150     160     170     180
orf120a.pep  DLFTLAWQLAANDAKLPPGLKITNGKKLYS VGLNKGATGKYSIGGVETE VVKYRVRGD
              |||
orf120-1      DLFTLAWQLAANDAKLPPGLKITNGKKLYS VGLNKGATGKYSIGGVETE VVKYRVRGD
              130     140     150     160     170     180

              190     200     210     220
orf120a.pep  DAVMYFFAPSLNNI PAQIGYTDGKTYTLKLKSVQINGQAAKFX
              |||
orf120-1      DAVMYFFAPSLNNI PAQIGYTDGKTYTLKLKSVQINGQAAKFX
              190     200     210     220

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF120 shows 97.8% identity over 184 aa overlap with a predicted ORF (ORF120ng) from *N.gonorrhoeae*:

```

55  orf120.pep  I PATMTFFERSGNAYKIVSTIKVPLYNIRFE 30
              |||
orf120ng      SAAILSAALPCAYAARLPQSAVLHYSGSYGI PATMTFFERSGNAYKIVSTIKVPLYNIRFE 69

60  orf120.pep  SGGTVVGNLHPTYYRDIRRGKLYAEAKFADGSVITYGKAGESKTEQSPKAMD LFTLAWQL 90
              |||
orf120ng      SGGTVVGNLHPAYYKDIRRGKLYAEAKFADGSVITYGKAGESKTEQSPKAMD LFTLAWQL 129

```

-434-

```

5  orf120.pep  AANDAKLPGLKITNGKKLYSVGGLNKAGTGRKYSIGGVETEVRVYRVRRGDDAVMYFFAP 150
    orf120ng  AANDAKLPGLKITNGKKLYSVGGLNKAGTGRKYSIGGVETEVRVYRVRRGDDVYFFAP 189

    orf120.pep  SLNNIPAQIGYTTDDGKTYTLKLKSVQINGQAAPK 184
    orf120ng  SLNNIPAQIGYTTDDGKTYTLKLKSVQINGQAAPK 223

```

The complete length ORF120ng nucleotide sequence <SEQ ID 779> is:

```

10  1  ATGATGAAGA  CTTTAAAAA  TATATTTTCC  GCCGCCATT  TGTCCGCCG
    51  CCTGCCGTGC  GGGTATGCGG  CAAGGCTACC  CCAATCCGCC  GTGCTGCACT
    101  ATTCCGGCAG  CTAACGCATT  CCGGCCACGA  TGACATTTGA  ACGCAGCGGC
    151  AATGCTTACA  AAATCGTTTC  GACGATTAAA  GTGCCGCTAT  ACAATATCCG
    201  TTTGCAATCC  GCGGGTACGG  TTGTGCGCAA  TACCGTGAC  CCTGCCTACT
    15  251  ATAAAGACAT  ACGCAGGGGC  AAACGTGTAT  CGGAAGCCAA  ATTGCGCGAC
    301  GGCAGCGTAA  CTAACGCCAA  AGCGGGCGAG  AGCAAAACCG  AGCAAAAGCC
    351  CAAGGCTATG  GATTGTGTCA  CGCTTGCCTG  GCAGTTGGCG  GCAAATGACG
    401  CGAAACTCCC  CCGGGTCTG  AAAATCACCA  ACGGCAAAAA  ACTTTATTC
    451  GTCCGCGGCC  TGAATAAGGC  GGGTACGGGA  AAATACAGCA  TAggCGGCGT
    20  501  GGAACCGGAA  TGCTGCATAT  ATCGGGTGG  GCGCGCGAC  GATACGCTAA
    551  CGTATTTCTT  CGCACCGTCC  CTGAACAAAT  TTCCGCGACA  AATCGGCTAT
    601  ACCGACGACG  GCAAAACCTA  TAGCTGGAAG  CTCAAATCG  TGCAGATCAA
    651  CGGACAGGCC  GCCAAACCGT  AA

```

This encodes a protein having amino acid sequence <SEQ ID 780>:

```

25  1  MKMTEFNIFS  AAILSAALPC  AYAARLPQSA  VLHYSGSYGI  PATMTFERSG
    51  NAYKIVSTIK  VFLYNIRFES  GGTVVGNTLH  PAYYKDIRRG  KLYAEAKFAD
    101  GSVTYGKAGE  SKTEQSPKAM  DLFTLAWQLA  ANDAKLPGL  KITNGKKLYS
    151  VGLLNKAGTG  KYSIGGVETE  VRVYRVRRGD  DTVTYFFAPS  LNNIPAQIGY
    201  TDDGKTYTLK  LKSVQINGQA  AKP*

```

30 In comparison with ORF120-1, ORF120ng shows 97.8% identity in 223 aa overlap:

```

    10      20      30      40      50      60
orf120-1.pep  MKMTEFNIFSAAILSAALPCAYAAGLPQSAVLHYSGSYGI PATMTFERSGNAYKIVSTIK
orf120ng      MKMTEFNIFSAAILSAALPCAYAARLPQSAVLHYSGSYGI PATMTFERSGNAYKIVSTIK
    10      20      30      40      50      60

    70      80      90      100     110     120
orf120-1.pep  VFLYNIRFESGGTVVGNTLHPAYYKDIRRGKLYAEAKFADGGSVTYKGAGE SKTEQSPKAM
orf120ng      VFLYNIRFESGGTVVGNTLHPAYYKDIRRGKLYAEAKFADGGSVTYKGAGE SKTEQSPKAM
    70      80      90      100     110     120

    130     140     150     160     170     180
orf120-1.pep  DLFTLAWQLAANDAKLPGLKITNGKKLYSVGGLNKAGTGRKYSIGGVETEVRVYRVRRGD
orf120ng      DLFTLAWQLAANDAKLPGLKITNGKKLYSVGGLNKAGTGRKYSIGGVETEVRVYRVRRGD
    130     140     150     160     170     180

    190     200     210     220
orf120-1.pep  DAVMYFFAPSLNNIPAQIGYTTDDGKTYTLKLKSVQINGQAAPKX
orf120ng      DTVTYFFAPSLNNIPAQIGYTTDDGKTYTLKLKSVQINGQAAPKX
    190     200     210     220

```

This analysis, including the presence of a putative leader sequence in the gonococcal protein suggests that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 93

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 781>:

1 ATGTATCGGA GGAAGGGCG GGGCATCAAG CCGTGGATGG GTGCGGGTGC
 51 .GCGTTTGGC GCCTTGGTCT GCGTGGTTT CCGCTCGGC GATACCTTGA
 101 CTCGGTTTGC GGTTCGGCGC GTGCTGGCGT ATGATTGGGA CCTTTGGTCT
 151 GAATGTTTGC AGAARAGGG TTTGAACCGT GCATCCGCTT CGATGCTGCT
 5 201 GATGGTGTTC TCCTTGATT TGTGTGGCG ATTATTGGTG ATTATCGTCC
 251 CTATGCTGGT CGGGCAGTTC AACAAATTGG CATCGCGCT GCCCAATTA
 301 ATCGGTTTGA TGCAGAACAC GGTCTGGCGC TGCTTGAAGA ATACATCGG
 351 CGGATATGTG GAAATCGATC AGGCATCTAT TATTGGCTGG CTTACGGCGC
 10 401 ATACGGGAGA GTTGAGCAAC GCGCTTAAAG CCGTGGTTCC CTTTGTGATG
 451 AGGCAGGGCG GCAATATT..

This corresponds to the amino acid sequence <SEQ ID 782; ORF121>:

1 MYRRKRGRIK PWMGAGXAFA ALVWLVFALG DTLTPFAVAA VLAYVLDPLV
 51 EWLQKKGLNR ASASMSVMVF SLILLALLL IIVPMLVGQF NNLASRLPQL
 101 IGFQMNTLLP WLKNTIGGYV EIDQASITAA LQAHGELSN ALKAWFPVIM
 151 RQGGNI..

Further work revealed the complete nucleotide sequence <SEQ ID 783>:

1 ATGTATCGGA GGAAGGGCG GGGCATCAAG CCGTGGATGG GTGCGGGTGC
 51 GCGTGGTTGC GCCTTGGTCT GCGTGGTTT CCGCTCGGC GATACCTTGA
 101 CTCGGTTTGC GGTTCGGCGC GTCCTGGCGT ATGATTGGGA CCTTTGGTCT
 151 GAATGTTTGC AGAARAGGG TTTGAACCGT GCATCCGCTT CGATGCTGCT
 20 201 GATGGTGTTC TCCTTGATT TGTGTGGCG ATTATTGGTG ATTATCGTCC
 251 CTATGCTGGT CGGGCAGTTC AACAAATTGG CATCGCGCT GCCCAATTA
 301 ATCGGTTTGA TGCAGAACAC GGTCTGGCGC TGCTTGAAGA ATACATCGG
 25 351 CGGATATGTG GAAATCGATC AGGCATCTAT TATTGGCTGG CTTACGGCGC
 401 ATACGGGAGA GTTGAGCAAC GCGCTTAAAG CGTGGTTTCC CTTTGTGATG
 451 AGGCAGGGCG GCAATATTGT CAGCAGTATC GGCACCTGCG TGCTGCTTCC
 501 CTTGCTGCTT TACTATTGCC TGCTGGATTG GCAGCGGTGG TCGTGCGGCA
 551 TTGCCAATCT GGTTCGGAGG CGTTTGGCG GCTCTTATAC GCGCATACA
 601 GGCATTTTGA ACGAGGTATT GGGGGAATT TTTGCGGGCG AGCTTCTGGT
 30 651 AATGCTGATT ATGGGCTTGG TTAAGGTTT GGGATTGGTG CTGCTCGGCG
 701 TGGATTGGGG TGTGCTCATC GGTATGCTG CCGGATATT GGTGTGTC
 751 CCTTATCTCG GGGCTTTTGA GCGATPCTG CTTGCGACG TCGCGCGCT
 801 GCTCCATCTC GGTCTCGGGA ACGCATCTAT ATCGGTTTGG CGCGTTTTT
 851 CCGTAGGACA GTTCTCGAA AGTTTTTCA TTACCGGGA AATCGTGGGA
 35 901 GACCGTATCG GCGTCTCGCC GTTTTGGGT ATCTTTTCC TGATGCGGTT
 951 CGGGCAGCTG ATGGCGTTTG TCGGAATGTT GCGGGGATGT CTTTGGCCG
 1001 CCGTAACCTT GGTCTTGCTT CGGAGGGCG TGCAGAAATA TTTTGGCCG
 1051 AGTTTTTACC GGGCAGGTA G

This corresponds to the amino acid sequence <SEQ ID 784; ORF121-1>:

1 MYRRKRGRIK PWMGAGXAFA ALVWLVFALG DTLTPFAVAA VLAYVLDPLV
 51 EWLQKKGLNR ASASMSVMVF SLILLALLL IIVPMLVGQF NNLASRLPQL
 101 IGFQMNTLLP WLKNTIGGYV EIDQASITAA LQAHGELSN ALKAWFPVIM
 151 RQGGNIVSSI GNILLPLLL YYFLLDQWR SCGIKLVPR RFAGAYTRIT
 20 201 INRNLVLEGF LRGLDVLML MSLVYGLSLV LVGLDSGFAT GMLAGLVLFV
 45 251 PVLGAFGLL LATVALLIQ GSWAGLSW AVPAVGDFE SEFTTKIVG
 301 DRIGLSPPWF IFSIMAFGL MSFVGLMAGL PLAAVTLVL RSGVQKIFAG
 351 SFYRGR*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

50 ORF121 shows 98.7% identity over a 156aa overlap with an ORF (ORF121a) from strain A of *N. meningitidis*:

	10	20	30	40	50	60
orf121.pep	MYRRKRGRIK	PWMGAGXAFA	ALVWLVFALG	DTLTPFAVAA	VLAYVLDPLV	EWLQKKGLNR
55 orf121a	MYRRKRGRIK	PWMGAGXAFA	ALVWLVFALG	DTLTPFAVAA	VLAYVLDPLV	EWLQKKGLNR
	70	80	90	100	110	120
orf121.pep	ASASMSVMVF	SLILLALLL	IIVPMLVGQF	NNLASRLPQL	IGFQMNTLLP	WLKNTIGGYV

-436-

5	orf121a	ASASMSVMVFSLILLALLLLIIVPMLVGQFNNLASRLPQLIGFMQNTLLPWLKNTIGGVV	70	80	90	100	110	120
10	orf121a.pep	EIDQASIIAWLQAHTGELSNALKAWFPVLMRQGGNI	130	140	150			
15	orf121a	EIDQASIIAWLQAHTGELSNALKAWFPVLMRQGGNI	130	140	150	160	170	180
20	orf121a	SOGIAKLVPRRFAGAYTRITGNLNEVLGEFLRGQLLVMLIMGLVYGLGLVVLGLDSGFAT	190	200	210	220	230	240

The complete length ORF121a nucleotide sequence <SEQ ID 785> is:

15	1	ATGTATCGGA	GGAAAGGCG	GGGCATCAAG	CCGTGGATGG	ATGCCGGTGC
	51	GGCGTTTGCC	GCCTTGGTCT	GGCTGGTTTT	CGCGCTCGGC	GATACTTTGA
20	101	CTCCGTTTGC	GGTTGCGGCG	GTGCTGGCGT	ATGTATTGGA	CCCTTTGGTC
	151	GAATGTTTGC	AGAAAAAGGG	TTTGAAACCGT	GCATCCGCTT	CGATGTCGTG
25	201	GATGGTGTTC	TCCTTGATTT	TGTTGTGGCG	ATTATTGTTG	ATTATTGTCC
	251	CTATGCTGGT	CGGCGAGTTC	AACAATTGGG	CATCGCGCT	GCCCAATTA
30	301	ATCGCTTTGA	TGCGAGAACG	GTCCTCGTGG	TGGTTGAAA	ATACATACGG
	351	CGGATATGTG	GAATTCGATC	AGGCATCTAT	TATTGCGCTG	CTTCAGGCGC
35	401	ATACGGGCGA	GTTCAGCAAC	CGCGTTAAGG	CGTGGTTTCC	CGTTTGTATG
	451	AGGCAGGCGC	GCAATATTGT	CAGCAGTATC	GGCAACCTGC	TGCTGCTTCC
40	501	CTTGCTGCTT	TACTATTTC	TGCTGGATTG	GCAGCGGTGG	TGCTGCGGCA
	551	TTGCCAAACT	GGTTCGAGG	CGTTTTCGCG	GTGCTTATAC	GGCATTATCA
45	601	GGCAATTGTA	ACGAGGTATT	GGGCGAATTT	TTGCGCGGGC	AGCTTCGTGT
	651	GATGCTGATT	ATGGGTTTGG	TTTACGCGCT	GGGGTTGGTG	CTGGTCGGGC
50	701	TGGATTTCGG	GTTTGCAATC	GGTATGGTTG	CGGTTATTTC	GGTTTGTGTT
	751	CCCTATTGGG	GCGCGTTTAC	AGGAATCGTG	CTGGCAACCG	TGCGCGCTTT
55	801	GCTCCAGTTC	GGTTCGTGGA	ACGCGCATCT	GGCTGTTTGG	GCGGTTTGTG
	851	CGCTAGGACA	GTTCCTCGAA	AGTTTTTTCA	TACGCGCAAA	AATCGTGGGA
60	901	GACCTATGTC	CGGCGTGGC	GTTTGGGTT	ATCTTTGCG	TGATGCGCTT
	951	CGGCGAGCTG	ATGGGCTTTG	TCGGAATGTT	GGCGGATATG	CCTTTGGCGG
65	1001	CCGTAACCTT	GGTCTTGCTT	CGCGAGGCGC	TGCAGAAATA	TTTTGCGCGC
	1051	AGTTTTTACC	GGGCGAGGTA	G		

This encodes a protein having amino acid sequence <SEQ ID 786>:

40	1	MYRRKGRGKIP	PWMDAGAAFA	ALVWLVFALG	DTLTPFAVAA	VLAYVLDPLV
	51	EWLQKKGLNR	ASASMSVMVF	SLILLALLLL	IIVPMLVGQF	NNLASRLPQL
45	101	IGFMQNTLLP	WLKNTIGGVV	EIDQASIIAW	LQAHTGELSN	ALKAWFPVLM
	151	RQGGNIIVSSI	GNLLLLPLLL	YFLLDQWRW	SCGIAKLVP	RFAGAYTRIT
50	201	GNLNEVLGEF	LRGQLLVMLI	MGLVYGLGLV	LVGLDSGFAT	GMVAGILVYV
	251	PYLGAFTGLL	LATVAALLQF	GSWNGILAVW	AVFAVGQFLE	SFFITPKIVG
55	301	DRIGLSPFWV	IFSLMAFGQL	MGFVGMHLAGL	PLAAVTLVLL	REGVQKYFAG
	351	SFYRGR*				

45 ORF121a and ORF121-1 show 99.2% identity in 356 aa overlap:

50	orf121a.pep	MYRRKGRGKIPWMDAGAAFAALVWLVLFDLDTLTPFAVAAVLAYVLDPLVEWLQKKGLNR	10	20	30	40	50	60
	orf121-1	MYRRKGRGKIPWMDAGAAFAALVWLVLFDLDTLTPFAVAAVLAYVLDPLVEWLQKKGLNR	10	20	30	40	50	60
55	orf121a.pep	ASASMSVMVFSLILLALLLLIIVPMLVGQFNNLASRLPQLIGFMQNTLLPWLKNTIGGVV	70	80	90	100	110	120
	orf121-1	ASASMSVMVFSLILLALLLLIIVPMLVGQFNNLASRLPQLIGFMQNTLLPWLKNTIGGVV	70	80	90	100	110	120
60	orf121a.pep	EIDQASIIAWLQAHTGELSNALKAWFPVLMRQGGNIIVSSIGNLLLLPLLLYYFLLDQWRW	130	140	150	160	170	180
	orf121-1	EIDQASIIAWLQAHTGELSNALKAWFPVLMRQGGNIIVSSIGNLLLLPLLLYYFLLDQWRW	130	140	150	160	170	180
65	orf121a.pep	SCGIAKLVPRRFAGAYTRITGNLNEVLGEFLRGQLLVMLIMGLVYGLGLVVLGLDSGFAT	190	200	210	220	230	240
	orf121-1	SCGIAKLVPRRFAGAYTRITGNLNEVLGEFLRGQLLVMLIMGLVYGLGLVVLGLDSGFAT	190	200	210	220	230	240

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

Homology with a predicted ORF from *N.gonorrhoeae*

ORF121 shows 97.4% identity over a 156 aa overlap with a predicted ORF (ORF121ng) from

N.gonorrhoeae:

20	orf121.pep	MYRRKGRGIGKPFPMGAGAAPAALVWLVYALGDTLPFAVAAVLAYVLDPLVWELQKKGINR	60
	orf121.ng	MYRRKGRGIGKPFPMGAGAAPAALVWLVYALGDTLPFAVAAVLAYVLDPLVWELQKKGINR	60
25	orf121.pep	ASASMSVMVFSLLILLALLLIIVPMIVGQFNNLASRLPOLIGFMONTLLPWLNKTIGGYV	120
	orf121.ng	ASASMSVMVFSLLILLALLLIIVPMIVGQFNNLASRLPOLIGFMONTLLPWLNKTIGGYV	120
	orf121.pep	EIDQASIIAWLQAHGTGLSNSALKAFPPVMKGGNI	156
30	orf121.ng	EIDQASIIAWFOAHGTGLSNSALKAFPPVMKGGNIIVSTIGNILLPLLLYYFLDWHRW	180

An ORF121ng nucleotide sequence <SEQ ID 787> was predicted to encode a protein having amino acid sequence <SEO ID 788>:

35

	MYRLKQGGRIK	PWMGGAAMFA	ALVVLVVAIG	DTITPFVAVA	VLAYVLDPVL
51	EYLRQKKGNL	ASASMSVGVN	SLIILLALLL	LIIVPMLVGQF	NILNASRLPGL
101	IQGMNTLLP	WLKNTIGGVV	EDQASITAW	FOARTGELSN	ALKAWFPVIM
151	KGGNIVSTI	SCGILLPELL	YYFLDLNHRH	SCGIPKLVP	RGAGAYTRT
201	GNNVWKQGF	LRGQLLGETE	RGAVVCRVGR	ECWEGGGARS	RPSDDGWPRW
251	GGN*				

Further work revealed the following gonococcal DNA sequence <SEO ID 789>:

40	1	ATGCTATCGGA	GAAGAAAGACG	GGGCATCAAG	CGCGCTGACG	GTGCGCGGCG
	51	GGCTGTATTC	GGCTTGGTCT	GGCTGTGTAT	CGGCGTCGGC	GACATCTATG
	101	CTCGGCTTGC	GGTGTGGCGG	GTGTGGAGGT	ATGTGTGTGA	CCCTTGGTCT
	151	GAATGGTGTTC	AGAAAAAAGG	GTGTGCGCTC	GCATCGCTGT	CGATGTCGTG
	201	GATGGTGTGTT	TCCTTGATTT	TTGTGTGGTC	ATATTATGTG	ATTATTTGTC
45	251	CTATGCTGTG	CGGCGAGTTC	TAATATTTGG	CACTCGCGCT	GCCCAATAT
	301	ATCGGTTTAT	TGCGAAGAAC	GCTGCTCGCG	TGGTGTGAAA	ATACATACAT
	351	CGGATATGTG	GAATCTGCAT	CGGATCTCAT	TATGCGGTT	TTTCAGGCGG
	401	ATACGAGTGC	TGTGTCGAAC	CGGCTTAAGG	CTGGTGTTCG	CGTTTTCAGT
	451	AACCGAGTGC	GGGCGGAGTC	GGGCGGAGTC	CGGCGGAGTC	CGGCGGAGTC
50	501	CTTGTGCTGT	TACTATTATC	TGCTGAGATG	GCAGCGGAGTC	TGCTGCGAGT
	551	TGCGCAAACT	GGTGTCCGAG	GGTGTCCGCG	GTGCTCATAC	GGCGCATACG
	601	GGTAATTTGA	ACGAGGATTT	GGGCGAGATC	TGTCGCGGTG	AGCCTTTCGT
	651	GGTGCCTGAT	ATTGGGCTTG	TATTCAGGTT	GGGATGTATG	CTAGTGGGAC
	701	TGGATTCGGG	ATTGGCCATC	GGTATGGAGT	CGCGTATATT	GGTGTTTGTC
55	751	CCCTATTGTC	GTGGGTTTAC	GGGATGTCGT	CTGCGCACTG	TGAGCACTAT
	801	GCTCCAGTTC	GGTGTGCTGA	ACGGATCTCT	GGCGTTTTGG	GGGGTTTTTG
	851	CGTGTGTGTC	ATTTCTTCGAA	AGTTTTCATC	TTAGCGCGGA	AATTTATGGA
	901	GACGCTATCA	GGTGTGGCGC	TTTGTGGGTT	GGTGTGGGTC	TTATGGGCTG
	951	CGGAGATGTC	ATGCTCTTTT	TGCGAGAGTC	CTTTTGGGTC	CTTTTGGGTC
	1001	GCTTAACCTT	GTCTTGCTGT	CGGAGGCGCG	CGCAGAAAT	TTTTTGCGCG
60	1051	AGTTTTTTACG	GGGCGAGGTA	G		

This corresponds to the amino acid sequence <SEQ ID 790; ORF12Ing-1>:

1	MYRRKGRGK	PWMGAGAAFA	ALVWL	VYALG	DTLTPFAVA	VLAYVLDPLV
51	EWLQKGLNR	ASASMSVMVF	SLTLLALLL	IIVPMLVGQF	NNLASRLPOL	
101	IGFMONTLLP	WLKNTIGGYV	ETDQASIIAW	FOAHTGELSN	ALKAWFPVLM	
151	KQGGNIVSSI	GNLLPLPLL	YFLLDQWRW	SCGIAKLVR	RFAGAYTRIT	
201	GNLNEVLGEF	LRGQLVLM	MGLVYGLGLM	LVGLDSGFAL	GMVAGILVVF	
251	PYLGAFTGLL	LATVAALLQF	GSWNGILAVM	AVFAVGQFLE	SFFITPKIVG	
301	DRIGLSPFWV	IFSLMAFGEI	MGFVGMLAGL	PLAAVTLLVL	REGAQKYFAG	
351	SFYRGR*					

10 ORF12Ing-1 and ORF12I-1 show 97.5% identity in 356 aa overlap:

	10	20	30	40	50	60
orf12i-1.pep	MYRRKGRGKIPWMGAGAAFAALVWL	VYALGDTLTPFAVA	VLAYVLDPLV	EWLQKGLNR		
orf12Ing-1	MYRRKGRGKIPWMGAGAAFAALVWL	VYALGDTLTPFAVA	VLAYVLDPLV	EWLQKGLNR		

	70	80	90	100	110	120
orf12i-1.pep	ASASMSVMVFSLILLALLL	IIVPMLVGQFNNLASRLPOL	IGFMONTLLPWLKNTIGGYV			
orf12Ing-1	ASASMSVMVFSLILLALLL	IIVPMLVGQFNNLASRLPOL	IGFMONTLLPWLKNTIGGYV			

	130	140	150	160	170	180
orf12i-1.pep	EIDQASIIAWLQHTGELSNALKAWFPVLM	RQGGNIVSSIGNLLPLLLYFLLDQWRW				
orf12Ing-1	EIDQASIIAWLQHTGELSNALKAWFPVLM	RQGGNIVSSIGNLLPLLLYFLLDQWRW				

	190	200	210	220	230	240
orf12i-1.pep	SCGIAKLVPRRFAGAYTRITGNLNEVLGE	FLRGQLVLMINGLVYGLGLMLVGLDSGFAL				
orf12Ing-1	SCGIAKLVPRRFAGAYTRITGNLNEVLGE	FLRGQLVLMINGLVYGLGLMLVGLDSGFAL				

	250	260	270	280	290	300
orf12i-1.pep	GMLAGILVFPYLGAFTGLLLATVAALLQF	GSWNGILAVMFAVAVGQFLESFFITPKIVG				
orf12Ing-1	GMLAGILVFPYLGAFTGLLLATVAALLQF	GSWNGILAVMFAVAVGQFLESFFITPKIVG				

	310	320	330	340	350	
orf12i-1.pep	DRIGLSPFWVIFSLMAFGQLMGFVGMLAGL	PLAAVTLLVLREGQKYFAGSFYRGRX				
orf12Ing-1	DRIGLSPFWVIFSLMAFGQLMGFVGMLAGL	PLAAVTLLVLREGQKYFAGSFYRGRX				

In addition, ORF12Ing-1 shows homology to a permease from *H. influenzae*:

	sp P43969 PERM_HAEIN PUTATIVE PERMEASE PERM HOMOLOG Length = 349
	Score = 69.9 bits (168), Expect = 2e-11
	Identities = 67/317 (21%), Positives = 120/317 (37%), Gaps = 7/317 (2%)
50	Query: 26 VYALGDTLTPFAVAALVYLDPLVEWL-QKKGILNRASASMSVMVFSXXXXXXXXXXVP 84
	+Y GD + P +A VL+Y+L+ +L Q R A++ + VP
	Sbjct: 32 IYFGDLIALPLLIALVLSYLLIPIINFLNQYLKCFRMLATILIFGSFIGLAAPFVLVLP 91
55	Query: 85 MLVGQFNNLASRLPOLIGFMONTLLPWLKNTIGGYVE-IDQASIIAWFOAHTGELSNALK 143
	ML Q +L S LP + N WL N Y E ID + + + F + + +
	Sbjct: 92 MLWNTISLLSDLPAMF-----NKSNEWLLNLPKNYPEDIDSMVDSIFNSVREKILGFGE 147
	Query: 144 AWFPVLMKQGGNIVSSIGNXXXXXXXXXXXXQWRWSCGIAKLVPRRFAGAYTRITGNL 203
60	+ + + N+VS D G+++P+ A+ R +
	Sbjct: 148 SAVKLSLASIMNLVSLGIYAFLVPLMMFFMLKDKSELLQGVSRFLPKNRLAFXRWK-EM 206
	Query: 204 NEVLGEFLRGQXXXXXXXXXXXXXXXXXDSGFAIGNVAGILVFPYXXXXXXXXXXXX 263
	+ + + G+ G+ V VP
65	Sbjct: 207 QQQISNYHGKLELLIIVTLITYIIFLIFGLNYPLLAFAGVLSVLVPYIGAVITPFA 266

Query: 264 XXXXXQFGSWGILAVWAVFAVGQFLESFFITPKIVGDRIGLSPFWVIFSLMAFGELMGF 323
 QFG + FAV Q L + + P + + + L P + I S + + F G L G F
 Sbjct: 267 LVALFQFGISPTFEWYIIIAFAVSQLLDGNLLVPYLFSEAVNLHPLIIISVLIFFGGLWGF 326

5 Query: 324 VGMLAGLFLAAVTLVLL 340
 G + + FLA + + +
 Sbjct: 327 WGVFFAIFLATLVKAVI 343

Based on this analysis, including the presence of a putative leader sequence and transmembrane domains in the two proteins, it is predicted that the proteins from *N.meningitidis* and
 10 *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 94

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 791>:

1 ..ACTGCTTTT CGGCGGCGCT GCGCTTGAGT CCATCATGAC TCGTCATATT
 15 51 TTTGTCTTTT GGGAAACCGT ATCAACAAC AGCCGCGCAT TTAACTATTT
 101 TTTGCACGTC CTGCGCGCGC CGTTCAAATG CGTACCAGCA ATACCGCGCG
 151 CTGCGCCTCT ATGCGCTTCCA TCCGCCGAG ATAGCCGAGT TTTTCGTTGG
 201 TTTTGCTTTT GATGTTGACG CACGAATGT CTATGCCCAA ATCGCGCGCG
 251 ATGTTGGCAC GATTTCGCG AATGTGCGG CGAGGTGTGG GTTTCGTGCG
 20 301 AATCACGGTC GTATCGACAT TGACCGCGCTG CCAACCTCGC GCCTGAACGC
 351 TTTGATACGC CGCACGCAAA AGGACGCGGC GTCCGCGATC TTTGAACCTT
 401 GCGCGCGTGT CGGGGAAATG GTCGCGGATA TCGCCCAAC CTGCGCGCAC
 451 GAGCAGCGCG TCGGTAACGG CGTGCGAGC CGCATGCCGA TCGGAGTGTC
 501 CGAGCAGCCC TTTTTCAAAT GGGATTTCAA CTCGCCCAAG TATCAG..

25 This corresponds to the amino acid sequence <SEQ ID 792; ORF122>:

1 ..TAFSAALRLS PSXLVIFLSF GKPYQQTAAI LFFCTSCPP RSNVQQYRR
 51 LRLYAFHPPE IAEFFVGFAG DVDARNVYAQ IGGDVGTHLR NVRRCGFLC
 101 NHGRIDIDRL PTLRLNALIR RTQKDAVRI FELCGGVGEM AADIAQTCT
 151 EQRVGVGVQV RIGIGVSEQP FFKWDFNSAK YQ..

30 Further work revealed the complete nucleotide sequence <SEQ ID 793>:

1 ATATCGTACT GGGCAAGCAG TTCGCCGAGT TTTTGGAA TAGATACCGC
 51 GCCTTTGATT TTTTGGCGC TCTTACCAA GGCTTCGATG AAAAAGTTGA
 101 TGGTCGAGCC GGTACCGATG CCGATATMT CATTTTCGGG TACGAATTCG
 151 ACTGCTTTT CGGCGGCGAT GCGCTTGAGT TCGTCTTGT TCGTCAATTT
 35 201 TTTGTCTTTT GGGAAACCGT ATCAACAAC AGCCGCGCAT TTAACTATTT
 251 TTTGCACGTC CTGCGCGCGC CGTTCAAATG CGTACCAGCA ATACCGCGCG
 301 CTGCGCCTCT ATGCGCTTCCA TCCGCCGAG ATAGCCGAGT TTTTCGTTGG
 351 TTTTGCTTTT GATGTTGACG CACGAATGT CTATGCCCAA ATCGCGCGCG
 401 ATGTTGGCAC GATTTCGCG AATGTGCGG CGAGGTGTGG GTTTCGTGCG
 40 451 AATCACGGTC GTATCGACAT TGACCGCGCTG CCAACCTCGC GCCTGAACGC
 501 TTTGATACGC CGCACGCAAA AGGACGCGGC GTCCGCGATC TTTGAACCTT
 551 GCGCGCGTGT CGGGGAAATG GTCGCGGATA TCGCCCAAC CTGCGCGCAC
 601 GAGCAGCGCG TCGGTAACGG CGTGCGAGC CGCATGCCGA TCGGAGTGTC
 45 651 CGAGCAGCCC TTTTTCAAAT GGGATTTCAA CTCGCCCAAG TATCAGCTTT
 701 CTGCGCTTGG TCGAGTTGGT GACATCGTAG CCTGTCCAG TACGAGTTGT
 751 CGTCATCGTT TGTGTTCTG A

This corresponds to the amino acid sequence <SEQ ID 794; ORF122-1>:

1 ISYWASSSPD FLEVDTAPLI FLELLPKASM KKIMVFPVPM PIYSFSGTNS
 51 TAFSAALRLS SSCVIFLSF GKPYQQTAAI LFFCTSCPP RSNVQQYRR
 50 101 LRLYAFHPPE IAEFFVGFAG DVDARNVYAQ IGGDVGTHLR NVRRCGFLC
 151 NHGRIDIDRL PTLRLNALIR RTQKDAVRI FELCGGVGEM AADIAQTCT
 201 EQRVGVGVQV RIGIGVSEQP FFKWDFNSAK YQLSAGQLV DIVALSDTDV
 251 RHRLCS*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF122 shows 94.0% identity over a 182aa overlap with an ORF (ORF122a) from strain A of *N. meningitidis*:

[illegible]

The complete length ORF122a nucleotide sequence <SEQ ID 795> is:

	1	TATATCATATT	GGGGCANGCAG	TTTCATCGGAT	TTTGTGGAG	TAGTACCCG
30	51	CGTCTTGATT	TTTGTGGCGG	AGGTCATCCAA	GGGTCGTGGA	AAAAAGTTGA
	101	CTGTGGTAACC	GTATACGACAT	CCGATGTATAT	CGTTTTCGGG	TACGAATATG
	151	ACTGTGNTT	CGGGCGGGAT	GGCGTTGAGT	TGCTCTGTG	TGCTCATAT
	201	TTTGTCTCTT	GGGAACCGT	ATCAACAAC	AGCGCCATC	TTAACATTT
	251	TTNNACGTC	GTCCGCGCG	CGTTCANACT	CTTACAGCA	CGTCCGTC
35	301	CTCGGACCTC	ATGCTCTTCA	TGCGCCGAG	ATAACGAGT	TTTCTGTGG
	351	TTTGGCTTT	GANGTTGACG	CAGGAATAT	CTATCGGCTA	ATCGCGGGG
	401	ATTTGGGCAC	GTAATTTGCG	AATATGGCG	CGGAGTTTGT	GGTTCGTGG
	451	AATACGGGTC	GCATTGCAGT	AGTCCGCTG	AGCAACCTG	GGCTGAACGC
	501	TTTGATACGG	AGCAGCGCAA	AGCGCGCGG	TGCTCGCATC	TTTGAACCTT
40	551	GGGCGGTGT	CGGGGAARTG	GCTGCGAAC	TGCGCCAAAC	CTCGCGCACC
	601	GAGACGACCG	TCGCTAACGG	CTGCGCAGC	CGCATCGCA	TGGAGTGTG
	651	GAGCAGCGCC	TTTTCARAAT	GGGATTTCAA	CGGCGATG	TATCAGATG
	701	CTGCTCTGTT	TCAGTTGGTG	CAGATCGTAG	CCCTGTTCGA	TACGAGTGT
	751	CGTATCGGT	TGTTGTCTG	A		

This encodes a protein having amino acid sequence <SEQ ID 796>:

45	1	ISYWSAMRLD	PLEVDFAPLI	FLPLLPKASM	KILMVEFVPM	PMYSFGSTNS
	51	TAFFSAMLRS	SCSVVILFSL	LKFKPYCTAAI	LTFEFTSCPPM	RNFYQCYRR
	101	LRIAYLAEPHE	TEFFYFAGFA	XVDARNVYQAQ	IGGDVGHHLR	NMRREKREK
	151	NHGRIDIDRL	PTLRNALRI	RQLCKDAVIR	FELCGGVGEM	AADIAQTCRT
	201	EQRVGNQVQQ	RIGIGVSEQP	FFKWFNSAK	YQLSAFGQLV	DIVALSDTDV
50	251	CRS*				

ORF122a and ORF122-1 show 96.9% identity in 256 aa overlap:

		10	20	30	40	50	60
55	orf122a.pep	ISYWASSSLDFLEVDTAFLIFLLPFLKASMKKLMVEVPMPMYSFSGTNTAFSAAMRLS					
	orf122-1	ISYWASSSLDFLEVDTAFLIFLLPFLKASMKKLMVEVPMPMYSFSGTNTAFSAAMRLS					
		10	20	30	40	50	60
60	orf122a.pep	SSCVVILFISGKPYQOTAAITFTFSCSPRSNFOYQVRRRLRLYAFHAPETITFFVGQFAP					
	orf122-1	SSCVVILFISGKPYQOTAAITFTFSCSPRSNFOYQVRRRLRLYAFHAPETITFFVGQFAP					

-441-

		70	80	90	100	110	120
		130	140	150	160	170	180
5	orf122a.pep	XVDARNVYAQIGGDVGT	HLRNMRRREFGFLCNHGRID	IDRLPTLRNLALIRRTQKDAAVRI			
	orf122-1	DVDARNVYAQIGGDVGT	HLRNMRRREFGFLCNHGRID	IDRLPTLRNLALIRRTQKDAAVRI			
		130	140	150	160	170	180
		190	200	210	220	230	240
10	orf122a.pep	FELCGGVGEMAAADIAQ	TCRTEQRVGVNGVQQRIGIGVSEQPF	FKWDFNSAKYQLSAFGQLV			
	orf122-1	FELCGGVGEMAAADIAQ	TCRTEQRVGVNGVQQRIGIGVSEQPF	FKWDFNSAKYQLSAFGQLV			
		190	200	210	220	230	240
		250					
15	orf122a.pep	DIVALSDTDVRHRLCSX					
	orf122-1	DIVALSDTDVRHRLCSX					
		250					
20							

Homology with a predicted ORF from *N. gonorrhoeae*

ORF122 shows 89.6% identity over a 182 aa overlap with a predicted ORF (ORF122ng) from

N. gonorrhoeae:

25	orf122.pep		TAFSAALRLSPXKLVIFLSFGKPYQQTAAI	30
	orf122ng	FLPLLPKASMKMLMVEPVM	PMPYSFSGTNSAFSAAMRLSSCVVIFLSFGKPYQQTAAI	80
	orf122.pep	LTFFCTSCPPRSNAYQQY	RRLLRYAFHPPEIAEEFVGFADVDARNVYAQIGGDVGT	90
	orf122ng	LTFFCTSWPPRSNAYQQY	RRLLRYAFHPPEIAEEFVGFADVDARNVYAQIGGDVGT	140
	orf122.pep	NVRRCEGFLCNHGRID	IDRLPTLRNLALIRRTQKDAAVRIFELCGGVGEMAAADIAQ	150
	orf122ng	NVRCFEGFLCNHGRID	IDHLPTLRNLALIRRTQKDAAVRIFELCGGVGEMAAADIAQ	200
35	orf122.pep	EQRVGNVGQQRIGIGVSEQ	PFKWDNFNSAKYQ	182
	orf122ng	EQRVGNVGQQRVGIRMP	EQPFKWDNFNSAKYQLSAFGQLVDIVALSDTDIRHRLCS	256

The complete length ORF122ng nucleotide sequence <SEQ ID 797> is:

40	1	ATGTCGTACC	GGGCAAGCAG	TTGCGCCGAT	TTT'TTGAGG	TTGAAACCGC
	51	GCCTTTGATT	TTTTTACCGC	TTTGCCCAAA	GGCTCGGATG	ARGAARTTGA
	101	tgcttgaacc	ggtaccgagt	ccgattgatt	cgttttccgg	tacgaatttcg
	151	ACTGCTTTT	CGCGCGCGAT	CGCcttgAgT	TGctcttgcg	TgTcATATT
	201	TTTAcctttt	gGGAaaccct	atcaAcaAAc	agccgcacat	TTAcATTTT
45	251	TTTGACGctc	ctggccgcgcg	cgttcaAATc	cgtagcaCAc	ataccgcgcg
	301	ctgcgcctCT	AtgcCTTCCA	TCGCGCCGAG	ATAGCCGAGT	TTTTCGTTGG
	351	TTTTGCTTTT	GATatTGACG	CACGAATAT	CGatacCCAa	atccggcgGCG
	401	ATGTITGGCAC	GCATITGGCG	AATGTGCGGT	CGAGTTTGG	GTTTCTGTGC
	451	AATCAGCGTC	GTATCGACAT	TGACCACCTG	CCAACCTCTG	GCCTGAACGC
50	501	TTTGATACGC	CGCACGCAAA	AGGACGCGCG	TGTCGCGATC	TTTGAACCTC
	551	GGCGCGGTGT	CGGGAAAATG	GCTGCGGATG	TGCGCCAAAC	CTCGCGCACC
	601	GAGCAGCGcg	tcggttaaCGG	CGTGCGAGAG	cgcgTcgGCA	TCCGAATGCC
	651	CGAGCAGCCC	TTTTTCAAA	GGGATTTCAA	CTCGCCCAAG	TATCAGCTTT
	701	CTGCCTTCGG	TCAATTGGTG	GACATCGTAG	CCCTGTGCGA	TACGATATT
55	751	CGTCATCGT	TGTGTTCTCG	A		

This encodes a protein having amino acid sequence <SEQ ID 798>:

	1	MSYRASSSPD	FLEVETAPLI	FLPLPKASM	KKLMVEPVM	PMPYSFSGTNS
	51	TAFSAAMRLS	SSCVVIFLSP	GKPYQQTAAI	LTFFCTSNPP	RSNPFYQYRR
60	101	LRLYAFHPPE	IAEEFVGFAD	IDARNIDTQ	IGGDVGT	HLRNVRCFGLC
	151	NHGRIDIDHL	PTLRNLALIR	RTQKDAVRI	FELCGGVGM	AADVAQTCRT
	201	EQRVGNVGQQ	RVGIRMP	EQPFKWDNFNSAK	YQLSAFGQLV	DIVALSDTDI
	251	RHRLCS*				

ORF122ng and ORF122-1 show 92.6% identity in 256 aa overlap:

		10	20	30	40	50	60
	orf122-1.pep	TSYWASSSPDFLEVDTAPLIFLPLPKASMKMLVPEVPMFIYSFSGTNSTAFSAAMRLS					
5	orf122ng	MSYRASSSPDFLEVETAPLIFLPLPKASMKMLVPEVPMFIYSFSGTNSTAFSAAMRLS					
		10	20	30	40	50	60
		70	80	90	100	110	120
10	orf122-1.pep	SSCVVIFLSFGKPYQQTAAILTFCTSCPPSRNAYQQYRRLRLYAFHPPEIAEFFVGF					
	orf122ng	SSCVVIFLSFGKPYQQTAAILTFCTSWPPSRNPYQQYRRLRLYAFHPPEIAEFFVGF					
		70	80	90	100	110	120
		130	140	150	160	170	180
15	orf122-1.pep	DVDARNVYAIGGDVGLHNRNVRREFGLCNHGRIDIDRLPTLRNALIRRTCKDAAVRI					
	orf122ng	DIDARNIDTQIGGDVGLHNRNVRCEFGFLCNHGRIDIDHLPRLNALIRRTCKDAAVRI					
		130	140	150	160	170	180
20	orf122-1.pep	FELCGGVGMAADIAQTCTEQRVGNVQQRIGIGVSEQPFKWDVNSAKYQLSAFGQLV					
	orf122ng	FELCGGVGMAADVAQTCTEQRVGNVQQRIGVIRMPQPFKWDVNSAKYQLSAFGQLV					
		190	200	210	220	230	240
25		190	200	210	220	230	240
		250					
	orf122-1.pep	DIVALSDTDVRRHRLCSX					
30	orf122ng	DIVALSDTDIRHRLCSX					
		250					

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 95

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 799>:

35	1	..GCCGGCGGCA GTGCGAACA CAITTCGCG CGTTTTGCGG AAACACCGGT
	51	CGCTGTCAGC GTTACCTGA TCGGCACGGT ACTTGCCGTC ATGCTGCCGC
	101	TTACCGAATA TGAATAATC CTGCTGCTTA TCGGCTCGGT ATTTGCGCGG
	151	ATGGGGCGGA TTTTGATTGC CGACTTTTTC GTCTTGAAC GGGGTGA

This corresponds to the amino acid sequence <SEQ ID 800; ORF125>:

40	1	..AGASANNISA RFAETPVAVS VTLIGTVLAV MLPVTEYENF LLLIGSVFAP
	51	MGGFDCRLFR LETA*

Further work revealed the complete nucleotide sequence <SEQ ID 801>:

	1	ATGTCGGGCA ATGCTCCTC TCCTTCATCT TCCTCGGCA TCGGGCTGAT
45	51	TTGGTTCGGC GGGCGGGTAT CGATTGCGA AATCAGCAG GGTACGCTGC
	101	TTGCGCCTTT GGGCTGGCAG CGGGCTCTGG CGGCTCTACT TTTGGGTCTAT
	151	GCGCTCGGCG GGGCGCTGTT TTTTGGGCG CGGTATATCG GGGCACTGAC
	201	CGGACGCGAG TCGATGAAA GGGTGCGCCT GTCGTTCCGG AAACGGGGTT
	251	CAGTGCTGTT TTCCGTGGCG AATATGCTCG AACTGGCCGG CTGGACGGCG
	301	GTGATGATTT AGCGGGGGC AACGCTCAGC TCCTCTTGG GCAAAGGTTT
	351	CTGGACGGCG GAATCTTTTG TCTGTGGTGG ATTGGCAAC GGGCGGCTGA
50	401	TTTGCTGTGG GCTGGTTTTC GGGCAGCGCA AATCAGGGCG GCTGAAACCC
	451	GTTCGATGC TGCTGATGCT GTTGCGGGTT CTGTGGCTGA GTGCGGAAGT
	501	CTTTTCCACG CGAGGCGAGCA CGGCGGCACA GGTTCAGAG GGCATGAGTT
	551	TGGAAACGGC AGTCGAGCTG TCCGCGCTGA TGCCGCTTTC CTGGCTCGCG
55	601	CTTGGCGGCG ACTACAGCGG CCACGGCGCG CGCCCGTTTG GGGCAACCTT
	651	GACGGCAACG CTGCGCTACA CGCTGACGGG CTGCTGGGATG TATGCTTTGG
	701	GTTTGGCAGC GGGCTGTGTC ACCGGAGAAA CCGACGTGGC AAAAATCTGT
	751	CTGGGCGCAG GTTTGGGTGC GGCAGGCATT TTGGCGGCTG TCCTCTCCAC

-443-

5
 801 CGTTACCACA ACGTTTCTCG ATGCCTATTC CGCCGGCGCG AGTGCGAACA
 851 ACATTTCGCG GCGTTTTCGCG GAAACACCGC TCGCTGTCGG CGTTACCCCTG
 901 ATCGGACGCG TACTTGCCGT CATGCTGCCG GTTACCGAAT ATGAAACCTT
 951 CCTGCTGCTT ATCGGCTCGG TATTTGCGCC GATGGCGGCG GTTTTGATTG
 1001 CCGACTTTTT CGTCTTGAAA CGGCGTGAGG AGATTGAAGG CTTTGACTTT
 1051 GCGGAGCTGG TTCTGTGGCT TCGGCGCTTC ATCTCTTACC GCTTCTGTCT
 1101 CTGTCCTGGC TGGGAAGACA GCATCGGTCT GACCGCGCCC GTAAAGTCTG
 1151 CCGTTGCAT TCGCACGTA TCGGTACGCC TTTTCTTTAA AAAACCCCA
 1201 TCTTTACAAA GGAACCGCTC ATGA

10 This corresponds to the amino acid sequence <SEQ ID 802; ORF125-1>:

1 MSGNASSPSS SSAIGLIWFG AAVSIAEIST GTLLAPLWQ RGLAALLGH
 51 AVGGALFFAA AYIGALTGRS SMESVRLSFG KRGSVLFSA NMLQLAGWTA
 101 VMIYAGATVS SALGKVLWDG ESFVWALAN GALIVLWLVF GARKTGGLKT
 151 VSMILLMLAV LWSARVEST AGSTAAQVSD GMSFGTAVEL SAVMPLSWLF
 201 LAADYTRHAR RPPAATLTAT LAYTLTGWM YALGLAALF TGETDVAKIL
 251 LAGLAGAAGI LAVVLSTVTT TFLDAYSAGA SANNISARFA ETPVAVGVTL
 301 IGTVLAVMLP VTEYENFLLI IGSVFAPMAA VLIADFFVLK RREEIEGDFD
 351 AGLVLWLAGF ILYRFLSSG WESSIGLTAP VMSAVAIATV SVRLFFFKTKQ
 401 SLQRNPS*

20 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF125 shows 76.5% identity over a 51aa overlap with an ORF (ORF125a) from strain A of *N.*

meningitidis:

25 orf125.pep 10 20 30
 AGASANNISARFAETPVAVSVTLTGLVAV
 orf125a KILLGAGLAGAAGILAVVLSTVTTTFLDAYSAGVSANNISAKLSEIPIAVAVVGTLLAV
 250 260 270 280 290 300
 30 orf125.pep 40 50 60
 MLPVTEYENFLLIGSVFAPMGGDFCRLFRLETAX
 :|||||:|||||:|||||:|||||:
 orf125a LLPVTEYENFLLIGSVFAPMAAVLIADFFVLKRREEIEG
 310 320 330 340

35 The ORF125a partial nucleotide sequence <SEQ ID 803> is:

1 ATGTCGGGCA ATGCCTCCTC TCNTTCATCT TCCGCCGCCA TCGGGCTGAT
 51 TTGGTTCGCG GCGGCGGTAT CGATTGCGGA AATCAGCAGC GGTACACTGC
 101 TTGGCGCTTT GGGCTGGCAG CGCGGCTGCG CNGCTCTGCT TTTGGGTGAT
 151 GCGGTCGGCG GCGCGCTGTT TTTTGGCGCG GCGTATATCT CGGCACTGAC
 40 201 CGGACNCANC TCGATGGAAA GCGTGCGCCT GTCGTTGCGG AAACGCGGTT
 251 CAGTGCTGTT TTGCGTGGCG AATATGCTGC AACTGCGCGG CTGGACGGCG
 301 GTGATGATT ACGGCGCGCG AACGCTCAGC TCGCTTTTGG GCAAAGTGTT
 351 GTGGGACGGC GATCTCTTTG TCTGGTGCGG ATTGGCAACG GCGGCTGA
 401 TTGCTGCTGT GCTGCTTTTC GCGCCACGCA AACACGGGCG GCTGAAAACC
 45 451 GTTTCGATGC TGCTGATGCT GTTGGCGGTT CTGCGCTGTA GTCCGCAANT
 501 NTTTTCCACG GCAGCGCAGCA CCGCGCGCANN GGTNNACAGC GGCAATGAGTT
 551 TCGGAACGCG AGTCGAGCTG TCCGCGGTNA TGCCGCTTTC TTGGCTGCCG
 601 CTGGCCGCGCG ACTACACGCG CCAAGCGGCG CCGCCGCTTG CGGCAACCCCT
 651 GAOGCGCAACG CTCGCTACCA CGCTGACGCG CTGCTGGATG TATGCTTTGG
 701 GTTTGGCAGC GCGCTTGTTT ACCGGAAGAA CCGACGCTGG AAAAATCTGT
 751 CTGGGCGCAG GTTTGGGTGC GCGACGCAIT TTGGCGGTGC TCGTCTGCAC
 801 CGTTACCACC ACTTTTCTCG ATGCTNACTC CGCCGCGGTA AGTGCCCAACA
 851 ATATTTCCCG CAACACTTTC GAAATACONA TCGCGCGTAT GCTGCGCGTT
 901 GTCGGCACAC TGCCTGCCGT CTTCTGCCCC GTTACCGAAT ATGAAACCTT
 951 CTGCTGCTT ATCGGTCGCG TATTTGCGCC GATGGCGGCG GTTTTGATTG
 1001 CCGACTTTTT CTTCTTGAAA CGGCGTGAGG AGATTGAAGG C.

This encodes a protein having the partial amino acid sequence <SEQ ID 804>:

1 MSGNASSXSS SSAIGLIWFG AAVSIAEIST GTLLAPLWQ RGLAALLGH
 51 AVGGALFFAA AYIGALTGXX SMESVRLSFG KRGSVLFSA NMLQLAGWTA

-444-

101 VMIYAGATVS SALGKVLWDG ESFVWVALAN GALIVLWLVF GARKTGGLKT
 151 VSMLLMLLAV LWLSAEXFST AGSTAAKXVD GMSFGTAVEL SAVMPLSWLP
 201 LAADYTRHAR RPFATLTAT LAYTLTGCMW YALGLAALF TGETDVAKIL
 251 LGAGLGAAGI LAVVLSVTVT TFLDAYSAGV SANNISAKLS SEPIAVAVAV
 301 VGTLLAVLLP VTEYENFLLI IGSVFAPMAA VLIADFFVLK RREEIEG..

ORF125a and ORF125-1 show 94.5% identity in 347 aa overlap:

		10	20	30	40	50	60
10	orf125a.pep	MSGNASSSSSSAIGLWFGAAVSIAEISTGTLLAPLQWQGLAALLGHAVGGALFPA					
	orf125-1	MSGNASSPSSSSAIGLWFGAAVSIAEISTGTLLAPLQWQGLAALLGHAVGGALFPA					
		10	20	30	40	50	60
15	orf125a.pep	AYIGALTGXSMESVRLSFGKRGSVLFSVANMLQLAGWTAVMIYAGATVSSALGKVLWDG					
	orf125-1	AYIGALTGRSSMESVRLSFGKRGSVLFSVANMLQLAGWTAVMIYAGATVSSALGKVLWDG					
		70	80	90	100	110	120
20	orf125a.pep	ESFVWVALANGALIVLWLVFGARKTGGLKTVSMLLMLLAVLWLSAEXFSTAGSTAAKXVD					
	orf125-1	ESFVWVALANGALIVLWLVFGARKTGGLKTVSMLLMLLAVLWLSAEXFSTAGSTAAKXVD					
		130	140	150	160	170	180
25	orf125a.pep	GMSFGTAVELSAVMPLSWLPLAADYTRHARRPFAATLTATLAYTLTGCMWYALGLAALF					
	orf125-1	GMSFGTAVELSAVMPLSWLPLAADYTRHARRPFAATLTATLAYTLTGCMWYALGLAALF					
		190	200	210	220	230	240
30	orf125a.pep	TGETDVAKILLGAGLGAAGILAVVLSVTVTTFDAYSAGVSANNISAKLSSEPIAVAVAV					
	orf125-1	TGETDVAKILLGAGLGAAGILAVVLSVTVTTFDAYSAGVSANNISAKLSSEPIAVAVAV					
		250	260	270	280	290	300
35	orf125a.pep	VGTLLAVLLPVTEYENFLLIIGSVFAPMAAVLIADFFVLKRREEIEG					
	orf125-1	IGTVLAVMLPVTEYENFLLIIGSVFAPMAAVLIADFFVLKRREEIEGDFAGVLVLWAGF					
		310	320	330	340	350	360

Homology with a predicted ORF from *N.gonorrhoeae*

45 ORF125 shows 86.2% identity over a 65aa overlap with a predicted ORF (ORF125ng) from *N.gonorrhoeae*:

	orf125.pep	AGASANNISARFAETPVAVSVTLIGTVLAV	30
	orf125ng	KILLGAGLGITGILAVVLSVTVTTFDLYTSAGASANNISARFAEIPVAVGTTLRTVLAV	308
50	orf125.pep	MLPVTEYENFLLIIGSVFAPM-GGFDCLRLFRLETA	64
	orf125ng	MLPVTEYKNFLLIRSVFGPMAGGDFCLRLFKTA	343

An ORF125ng nucleotide sequence <SEQ ID 805> was predicted to encode a protein having amino acid sequence <SEQ ID 806>:

1 MSGNASSPSS SAAIGLWFG AAVSIAEIST GTLLAPLQWQ RGLAALLLGH
 51 AVGGALFPA AYIGALTGRS SMESVRLSFG KCGSVLFSVA NMLQLAGWTA
 101 VMIYAGATVS SALGKVLWDG ESFVWVALAN GALIVLWLVF GARKTGGLKT
 151 VSMLLMLLAV LWLSVEVFAS SGTNAAPAVS DGMTFGTAVE LSAVMPLSWL
 201 PLAADYTRQA RRPFAATLTA TLAYTLTGCM MYALGLAALF FTGETDVAKI
 60 251 LLGAGLGITG ILAVVLSVTVT TFLDAYSAGV SANNISARF AEIPVAVGT
 301 LIRTVLAVML PVTEYKNFLL LIRSVFGPMA GGFDCLRLCL KTA*

Further work revealed the following gonococcal DNA sequence <SEQ ID 807>:

```

1  ATGTCGGGCA  ATGCTCTCTC  TCTTTCATCT  TCCGCGCCCA  TCGGGCTGGT
51  TTGGTTCGCG  GCGCGCGTAT  CGATTGCCGA  AATCAGCAGC  GGTAAGCTGC
101  TCGCCOCCCTT  GGGCTGSCAG  CCGGCTCTGG  CGGCCCTGCT  TTTGGGTGAT
151  GCGCTCGGGC  GCGCGCTGTT  TTTTCCGGCG  GCGTATCGAC  GCGCACTGAC
201  CGGACGCGAG  TCGATGGAAA  GTGTGCGCCT  GTCGTCGCG  AATGCGGTTT
251  CAGTGCTGTT  TTCCGTGGCG  AATATGCTGC  AACTGCGCGG  CTGGACGGCG
301  GTGATGATTT  ACGTCGGCGC  AACGGTCAGC  TCCGCTTTGG  GCAAGTGTTT
351  GTGGGACGCG  GAATCCCTTG  TCTGTGGGCG  ATTGGCAAAAC  GGCGCACTGA
401  TCGTGTCTGT  GCTGCTTTTC  GCGCAACGCA  GAACGGCGCG  GCTGAAAACC
451  GTTTCGATGC  TGCTGATGCT  GCTTGGCGTG  TTGTGGTTGA  GCGTCGAAGT
501  GTTTCGCTTG  TCCGSCACAA  ACGCGCGGCC  CGCGCTTTCA  GACGCGATGA
551  CTTTGGGAC  GCGAGTCGAA  CTGTGCGCG  TCATGCGGCT  TTCTCGGCTG
601  CCGCTGGCG  CCGCATACAC  GCGSCAAGCA  CCGCGCCGCT  TTGCGGCAAC
651  CCGTACGGCA  ACGCTCGCCT  ATACGCTGAC  GCGCTGCTGG  ATGTATGCTT
701  TGGGTTTGGC  GCGGCTCTGT  TTTACCGGAG  AAACCGACGT  GCGGAAAATC
751  CTGTTGGGCG  GCGGCTTGGG  CATAACGGCG  ATTCTGGCAG  TCGTCTCTCT
801  CACCGTTACC  ACAACGTTTC  TCGATACCTA  TTCCGCGCGG  GCGAGTCGGA
851  ACAACATTTC  CCGCGGTTTT  GCGGAAATAC  CCGTCGCTGT  CCGCGTTACC
901  CTGATCGGCA  CCGTGCTTGC  CGTCACTGCT  CCGCTTACCG  AATATAAAAA
951  CTTCCTGCTG  CTTATCGGCT  CGGTATTGCG  GCGCATGGCG  GCGGTTTTGA
1001  TTGCGCACTT  TTTGCTCTTA  AAACGGCGTG  AGGAGATTGA  AGGCTTTGAC
1051  TTTGCGGAC  TGGTCTGTGT  GGTGGCAGCG  TTCTCTCTCT  ACCGCTTCCT
1101  GCTCTCGTCC  GGTGGGAAA  GCGACATCGC  TCTGACCGCC  CCGGTAAATG
1151  TCGCGTTGCG  CATTGCGACC  GTATCGGTAC  GCGTTTCTTT  TAAAAAAACC
1201  CATCTTTAC  AAGGAACCC  GTCATGA

```

This corresponds to the amino acid sequence <SEQ ID 808; ORF125ng-1>:

```

1  MSGNASSPSS  SAAIGLVWFG  AAVSIAEIST  GTLLAPLQWQ  RGLAALLLGH
51  AVGGALFFRA  AYGALTGRS  SMESVRLSPG  KCGSVLFSVA  NMLQLAGWTA
101  VMIYVGATVS  SALGKVLWDG  ESEVVMALAN  GALVILVLMVF  GARRTGGLKT
151  VSMLLMLLAV  LNLGVEVFAS  SGTNAAPAVS  DGMTFTGTA  LSVAAMPLSLW
201  PLAADYTRQA  RRPFAATLTA  TLAYLTGCM  MYALGLAAAL  FTGETDVAKI
251  LLGAGLGITG  ILAVVLSVT  TTFELDTYSAG  ASANNISARF  AEIPVAVGVT
301  LIGTVLAVML  PUTEYKNEFL  LIGSVFAPMA  AVLTADEFVL  KRREEIEGFD
351  FAGVLVLAG  FILYRFLSS  GWESSIGLTA  PVMSAVAIAT  VSVLEFFPKT
401  QSLQRNPS*

```

ORF125ng-1 and ORF125-1 show 95.1% identity in 408 aa overlap:

```

10      20      30      40      50      60
orf125-1.pep  MSGNASSPSSSSAIGLVWFGAAVSIAEISTGTLLAPLQWQRLAALLLGHAVGGALFFRA
40      |||||
orf125ng-1    MSGNASSPSSSSAIGLVWFGAAVSIAEISTGTLLAPLQWQRLAALLLGHAVGGALFFRA
10      20      30      40      50      60
45      70      80      90      100     110     120
orf125-1.pep  AYIGALTGRSSMESVRLSFGKRGSVLFSVANMLQLAGWTAVMIYAGATVSSALGKVLWDG
45      |||||
orf125ng-1    AYIGALTGRSSMESVRLSFGKRGSVLFSVANMLQLAGWTAVMIYAGATVSSALGKVLWDG
70      80      90      100     110     120
50      130     140     150     160     170     179
orf125-1.pep  ESFVWWALANGALIVLWLVFGARKTGGLKTVSMLMLLAVLWLSAEVFTAGSTAAQ-VS
45      |||||
orf125ng-1    ESFVWWALANGALIVLWLVFGARKTGGLKTVSMLMLLAVLWLSAEVFTAGSTNAAPAVS
130     140     150     160     170     180
55      180     190     200     210     220     230     239
orf125-1.pep  DGMSTFGTAVLSAVMPLSWPLAADYTRHARRPFAATLTATLAYLTGCGWYALGLAAAL
45      |||||
orf125ng-1    DGMTFGTAVLSAVMPLSWPLAADYTRQARRPFAATLTATLAYLTGCGWYALGLAAAL
190     200     210     220     230     240
60      240     250     260     270     280     290     299
orf125-1.pep  FTGETDVAKILLGAGLGAAGILAVVLSVTVTTFELDTYSAGASANNISARFAETPVAVGVT
45      |||||
65      |||||
orf125ng-1    FTGETDVAKILLGAGLGSILAVVLSVTVTTFELDTYSAGASANNISARFAETPVAVGVT
250     260     270     280     290     300

```

-446-

		300	310	320	330	340	350	359
5	orf125-1.pep	LIGTVLAVMLPVTEYENFLLIGSVFAPMAAVLIADFFVLKRREEIEGDFAGLVMLAG						
	orf125ng-1	LIGTVLAVMLPVTEYKNFLLIGSVFAPMAAVLIADFFVLKRREEIEGDFAGLVMLAG						
		310	320	330	340	350	360	
10	orf125-1.pep	FILYRFLSSGWESSIGLTAPVMSAVAIATVSVRLFKKTSLQRNPSX						
	orf125ng-1	FILYRFLSSGWESSIGLTAPVMSAVAIATVSVRLFKKTSLQRNPSX						
		370	380	390	400			

Based on this analysis, including the presence of putative leader sequence and transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 96

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 809>:

20	1	ATGACCCGTA	TGCGCATCCT	CGCGCGGGCG	CTCTCGGGAA	GGCTGACCGC
	51	GTTCGAGCTT	GCAGAACAAAG	GTATACAGAT	TGCACTTTTC	GATAAAGGCT
	101	GCGCGCGGGG	CGAACACGCC	GCGCGCTATG	TGCGCGCGCG	CATGCTCGCG
	151	CCTGCGCGGG	AACGCTGCA	AGCCACGCC	GAATGGTCA	GCGTGGCGAG
	201	GCAGAGCATC	CCGCTTTGGC	CGCGCATCCG	ATGCGGCTCG	AACACGCACA
	251	CGATGATGCA	GGAAACCGCG	AGCCTGATTG	TATGGCACGG	GCAGGACAAG
25	301	CCATTATCCA	GCGAGTTGCT	CGCGCATCTC	AAACGCGCGG	CGGTACGGA
	351	TGACGAAATC	GTCCGTTGGC	GCGCGACGCA	CATCGCGGAA	CGCGAACCGC
	401	AACTCGCGGG	ACGTTTTTAA	GACGCGCATC	ACCTGCGCAG	CGAAGCCAG
	451	CTCGACGGCG	GGCAATATATA	GTCTGCACTT	GCGGACGCTT	TGGACGAAC
	501	GAACGTCCCC	TGCCATTGGG	AACACGAATG	CGTCCCGGAA	GCGTGCAAG

30 This corresponds to the amino acid sequence <SEQ ID 810; ORF126>:

	1	MTRIALILGGG	LSGRLLTALQL	AEQGYQIALF	DKSCRRGEHA	AAYVAAAMLA
	51	PAAXTVEATP	EVVRLGRQSI	PLWRGIRCLR	NHTMFMQENG	SLIVVHGQDK
	101	PLSSEFVRHL	KRGGCTDDEL	VWRMRADIAE	REPGLGRFX	DGIYLPTEQX
	151	LDGRQLXSAL	ADALDELNVF	CHWEHCVEPE	ACK...	

35 Further work revealed the complete nucleotide sequence <SEQ ID 811>:

	1	ATGACCCGTA	TGCGCATCCT	CGCGCGGGCG	CTCTCGGGAA	GGCTGACCGC
	51	GTTCGAGCTT	GCAGAACAAAG	GTATACAGAT	TGCACTTTTC	GATAAAGGCT
	101	GCGCGCGGGG	CGAACACGCC	GCGCGCTATG	TGCGCGCGCG	CATGCTCGCG
	151	CCTGCGCGGG	AACGCTGCA	AGCCACGCC	GAATGGTCA	GCGTGGCGAG
40	201	GCAGAGCATC	CCGCTTTGGC	CGCGCATCCG	ATGCGGCTCG	AACACGCACA
	251	CGATGATGCA	GGAAACCGCG	AGCCTGATTG	TGTGGCACGG	GCAGGACAAG
	301	CCATTATCCA	GCGAGTTGCT	CGCGCATCTC	AAACGCGCGG	CGGTAGCGGA
	351	TGACGAAATC	GTCCGTTGGC	GCGCGACGCA	CATCGCGGAA	CGCGAACCGC
	401	AACTCGCGGG	ACGTTTTTCA	GACGCGCATC	ACCTGCGCAG	CGAAGCCAG
45	451	CTCGACGGCG	GGCAATATAT	GTCTGCACTT	GCGGACGCTT	TGGACGAAC
	501	GAACGTCCCC	TGCCATTGGG	AACACGAATG	CGTCCCGGAA	GCGTGCAAG
	551	CCCAATACGA	CTGCGTATAT	GACTSCGCGG	GCTACGCGCG	AAAACCGCGG
	601	TGCAACCAAT	CCCGGACGCA	CACCAACACC	CTGCGCGGCA	TACGCGCGGA
	651	AGTGCGCGGG	GTTTACACAC	CCGAATACAC	GCTCAACCGC	CCGTGCGCTC
	701	TGCTCATCCG	CGGTATACCG	CTCTACATCG	CCCGGAAAGA	AAACACAGTC
	751	TTGCTCATCG	GCGCGAACCA	AATCGAAGAG	GAAAGCCAGG	CCCGCGCGAG
	801	CGTGCGGTCA	GCGTTGGAAC	TCTTGTCCCG	ACTCTATCGC	ATCCACCGCG
	851	CCTTCGCGGA	AGCGGACATC	CTCGAATATG	CCACCGGCGC	GCGGCCCGAG
	901	CTCAACCAAC	ACAACCCGCA	AATCGGTTAC	AACCGCGCGC	GACGCTGTAT
55	951	TGAATCAAC	GGCTTTTTCG	GCCACGGTTT	CATGATCTCC	CCGCGCGTAA
	1001	CGCGCGCGCG	CGCCGAGTAT	GCAATGGCAC	TGTTTGAGCG	AAAAGACGCG
	1051	CCGGAACGCG	ATAAAGAAGG	CGGTTTGGCG	TATATCCGAA	GACAAGATTA
	1101	A				

This corresponds to the amino acid sequence <SEQ ID 812; ORF126-1>:

```

1  MTRIALGGG LSGRLTALQL AEQGYQIALF DKGCRERGEHA AAYVAAAMLA
51 PAEAWEATP EVVRLGRQSI PLWRGIRCLR NTHTMQENG SLIVWHGQDK
101 PLSSEFVRHL KRGGVADDEI VWRADDIAE REPQLGGRFS DGILYLPTEGQ
151 LDGRQILSAL ADALDELNVP CHWEHECVPE GLQAGYDWLI DCRGYGAKTA
201 WNQSPXHTST LRGRIGEVAR VYTPETILNR PVRLHFRYP LYIAPKENHV
251 FVIGATQIES ESQAPASVRS GLELLSALYA IHFAFGEDI LEIATGLRPT
301 LNHNPFEIRY NRARRLIEIN GLFRHGFMS PAVTAAARL AVALFDGKDA
351 PERDKESGLA YIRROD*

```

10 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF126 shows 90.0% identity over a 180aa overlap with an ORF (ORF126a) from strain A of *N.*

meningitidis:

```

15  orf126.pep  MTRIALGGGLSGRLTALQLAEQGYQIALFDKGCRRGEHAAYVAAAMLAPAAKTEATP
      orf126a  MTRIALGGGLSGRLTALQLAEQGYQIALFDKGCRRGEHAAYVAAAMLAPAAEAVEATP
      10      20      30      40      50      60

20  orf126.pep  EVVRLGRQSIPLWRGIRCLRNLNTHTMQENGSLIVWHGQDKPLSSEFVRHLKRGXKTDDEI
      orf126a  EVVRLGRQSIPLWRGIRCLRNLNTHTMQENGSLIVWHGQDKPLSSEFVRHLKRGVADDXI
      70      80      90      100     110     120

25  orf126.pep  VWRADDIAEREPOLGGRFXDGIYLPTEXQLDGRQLXSALADALDELNVPCHWEHECVPE
      orf126a  VWRADDIAEREPOLGGRFXDGIYLPTEXQLDGRQLXSALADALDELNVPCHWEHECAPE
      130     140     150     160     170     180

30  orf126a  VWRADDIAEREPOLGGRFXDGIYLPTEXQLDGRQLXSALADALDELNVPCHWEHECAPE
      130     140     150     160     170     180

```

The complete length ORF126a nucleotide sequence <SEQ ID 813> is:

```

1  ATGACCCGTA TGGCCATCCT CGGCGCGGGC CTCTCNCGAA GGCTGACCGC
51 ACTGCAGCTT GCAGAACCAAG GTTATCAGAT TGCACTTTTC GATAAAGGCT
101 GCGCGCGGGG CGAACACCGCC GTCGCCATAT TTGCGCGCCG CATGCTCCGC
151 CTGCGCGGGG AAGCGGTGCA AGCCACGCTT GAATGAGTCA GCGTGGGACG
201 GCAGACATCT CCGCTTTGGC GCGCGCGGGC ATGCCATCTG AAAACGGCTG
251 CCATGATGCA NGAAACCGGC AGCCTGATTG TGTGGCAGCG CGAGACAAA
301 CTTTATCCCA ACGAGTTTCT CGGCCATCTC AAACGCGGGC GCGTAGCGGA
351 TGACNAAATC GTCCGTTGGC GCGCGGACGA CATGCGCGAA CGCGAACCGC
401 AACTCGGCGG ACGTTTTTCA GACGCGATCT AACTGCGGAC CGAAGGCGAG
451 CTCGACGGGC GGCAATATT GTCTGCATTT GCGGACGCTT TGGACGAAT
501 GAACGTCGCC TGCCATTGGG AACACGAATG TGCCCGCGAA GACTTGCAAG
551 CCCAATACGA CTGGCTGATC GACTGCGCGG GTACGCGGCG AAAACCGCGG
601 TGGAAACCAAT CCGCCGAGNA NACCGACACC CTGCGCGGCA TACGCGGCGA
651 AGTGGCGGGG GTTTACACAC CGAATATCAC GGTCAACGCG CCGCTGCGGC
701 TGCTACACCC GCGCTATCCG CTNATCATCC CCGCGAAGAA AAACNCGCT
751 TTGCTCATCG CGCGACCCCA AATCGAAAGC GAAGACCAAG CACTGCGGAG
801 CGTGCCTGCC GGGCTGGGAC TCTATATCCC ACTCTATGCC GTTACCGGCG
851 CTTTGGGCGA AGCGGATCT CTCGACATCG CACACGCGCT CGCGGCGGCG
901 CTCATCACCA ACAACCCGGA AATCGGTTAC AACCGCGGCC GACGCTGAT
951 TGAATCAAC GGCCTTTTCC GCCACGGTTT CATGATCTCC CCGCCGTA
1001 CCGCGCGCGC CGTCGATTG GCACTGGCAC TGTTTGACGG AAAAGANGCG
1051 CCGGACCGCG ATGAAGAAAG CGGTTTGGCG TATATCGGAA GACAGATTA
1101 A

```

55 This encodes a protein having amino acid sequence <SEQ ID 814>:

```

1  MTRIALGGG LSGRLTALQL AEQGYQIALF DKGCRERGEHA AAYVAAAMLA
51 PAEAWEATP EVVRLGRQSI PLWRGIRCLR PLFAMKXENG SLIVWHGQDK
101 PLSNEFVRHL KRGGVADDEI VWRADDIAE REPQLGGRFS DGILYLPTEGQ
151 LDGRQILSAL ADALDELNVP CHWEHECAPE DLQAGYDWLI DCRGYGAKTA
201 WNQSPXHTST LRGRIGEVAR VYTPETILNR PVRLHFRYP LYIAPKENHV

```

251 FVIGATQIES ESQAPASVRS GLELLSALYA VHFAPGEADI LEIATGLRPT
 301 LNHHNFEIYRNRARLIEIN GLFRHGFMIS PAVTAAAVRL AVALFDGKXA
 351 FERDEESGLA YIRRQD*

ORF126a and ORF126-1 show 95.4% identity in 366 aa overlap:

5	orf126a.pep	10	20	30	40	50	60
		MTRIAILGGGLSGRLTALQLAEQGYQIALFDKGCRRGHEHAAAYVAAAMLAPAAEAVEATP					
	orf126-1	MTRIAILGGGLSGRLTALQLAEQGYQIALFDKGCRRGHEHAAAYVAAAMLAPAAEAVEATP					
10	orf126a.pep	70	80	90	100	110	120
		EVVRLGRQSIPLWRGIRCLNTHMTMCENGSLIVWHGQDKPLSSEFVRLHKGXVADDEI					
	orf126-1	EVVRLGRQSIPLWRGIRCLNTHMTMCENGSLIVWHGQDKPLSSEFVRLHKGXVADDEI					
15	orf126a.pep	130	140	150	160	170	180
		VNRADDAIEAREPQLGGRFSDGIYLPTEGQLDGRQLSALADALDELNPCHNEHECVPE					
	orf126-1	VNRADDAIEAREPQLGGRFSDGIYLPTEGQLDGRQLSALADALDELNPCHNEHECVPE					
20	orf126a.pep	190	200	210	220	230	240
		DLQAQYDWLIDCRGYGAKTAWNQSPXXTSLRGIRGEVARVYTPETILNRPVRLHPRYP					
	orf126-1	GLQAQYDWLIDCRGYGAKTAWNQSPXXTSLRGIRGEVARVYTPETILNRPVRLHPRYP					
25	orf126a.pep	250	260	270	280	290	300
		LYIAPKENXVVFVIGATQIESQAPASVRSGLELLSALYA VHFAPGEADILEIATGLRPT					
	orf126-1	LYIAPKENXVVFVIGATQIESQAPASVRSGLELLSALYA VHFAPGEADILEIATGLRPT					
30	orf126a.pep	310	320	330	340	350	360
		LNHHNFEIYRNRARLIEINGLFRHGFMISPAVTAARLAVALFDGKXAPERDEESGLA					
	orf126-1	LNHHNFEIYRNRARLIEINGLFRHGFMISPAVTAARLAVALFDGKXAPERDEESGLA					
35	orf126a.pep	310	320	330	340	350	360
		LNHHNFEIYRNRARLIEINGLFRHGFMISPAVTAARLAVALFDGKXAPERDEESGLA					
	orf126-1	LNHHNFEIYRNRARLIEINGLFRHGFMISPAVTAARLAVALFDGKXAPERDEESGLA					
40	orf126a.pep	YIRRQDX					
	orf126-1	YIRRQDX					

Homology with a predicted ORF from *N.gonorrhoeae*

ORF126 shows 90% identity over a 180 aa overlap with a predicted ORF (ORF126ng) from *N.gonorrhoeae*:

50	orf126.pep	MTRIAILGGGLSGRLTALQLAEQGYQIALFDKGCRRGHEHAAAYVAAAMLAPAAEAVEATP	60
	orf126ng	MTRIAVLGGGLSGRLTALQLAEQGYQIELFDKGTQCGEHAAYVAAAMLAPAAEAVEATP	60
	orf126.pep	EVVRLGRQSIPLWRGIRCLNTHMTMCENGSLIVWHGQDKPLSSEFVRLHKGXVADDEI	120
55	orf126ng	EVIRLGRQSIPLWRGIRCLNTHMTMCENGSLIVWHGQDKPLSSEFVRLHKGXVADDEI	120
	orf126.pep	VNRADDAIEAREPQLGGRFSDGIYLPTEGQLDGRQLSALADALDELNPCHNEHECVPE	180
	orf126ng	VNRADDAIEAREPQLGGRFSDGIYLPTEGQLDGRQLSALADALDELNPCHNEHECAQ	180

60 An ORF126ng nucleotide sequence <SEQ ID 815> was predicted to encode a protein having amino acid sequence <SEQ ID 816>:

1 MTRIAVLGGGLSGRLTALQLAEQGYQIELFDKGTQCGEHAAYVAAAMLAPAAEAVEATP

5
 51 PAASEAVEATP EVIRLGRQSI PLWRGIRCLR NLTMQENG SLIVWHGQDK
 101 PLSSEFVRHL KRGVADDEI VWRADSEIAE REPOLGGRFS DGIYLPTEGG
 151 LDGRQILSAL ADALDELNVF CHEHECAPQ DLQAQYDWVI DCRGYGAKTA
 201 WNQSPHSTST LRIGRGEVRS FTRPKSRSTA PCACCTRAIR STSPRKTTS
 251 SSSARPKSA KAKPPPAYVF GWNYSRPMSP STPPSAKPTS SKWRPGLRPT
 301 LNHNHPEIRY SRERRLIEIN GLFRHGFMS PAVTAAAVRL AVALFDGKDA
 351 PERDEESGLA YIGRQD*

Further work revealed the following gonococcal DNA sequence <SEQ ID 817>:

10
 1 ATGACCCGTA TGGCGTCTCT CGGAGGCGGG CTTTCGGGAA GGCTGACCGC
 51 ATTGCAGCTT CGACACACAG GTTATCAGAT TGAACCTTTT GACAGGGGCA
 101 CCGGCCAAGG CGAACACGCG CGCGCGCTATG TTGCGCGCGC GATGCTCGCG
 151 CCTGGCGCGG AAGCGGTGCA GGCACGCGCC GAAGTCATCA GGCTGGGCG
 201 CGAGAGCATT CCGCTTTGGC GGGCGATCGC ATGCGGCTGT AACACGCTCA
 251 CGATGATGCA GGAACACGCG AGCGTGATTG TGTGGCGCGG GCAGGACAG
 15
 301 CCATTATCCA GCGAGTTCTG CCGCGCATCT AACCGCGCGG GCGTAGCGGA
 351 TGACGAAATC GTCGCTTGGC GCGCGCATGA AATCGCGGAA CGCGAACCGC
 401 AACTCGGCGG ACGTTTTTCA GACGCGCATCT ACCTGCGGAC CGAAGCCGAG
 451 CTCGACGGGC GGCATAATTT GTCTGCACTT GCGCAAGCTT TGGACGAAT
 501 GAACGTCCTT TGGCATTTGG AACACGAATG CGCGCCCGCA GACCTGCAAG
 551 CCAATAACGA CTGGGTAAATC GACTGCGCGG GCTACGGCGG GAAACCGCG
 601 TGAACCAAT CCGCGAGACA CACGAGCGCC TTGCGCGGCA TAGCGCGGCA
 651 AGTGGCGCGG GTTATCAGCG CGGAACACAC GCTGACCGC CCGCTGCGCG
 701 TCGTGCACCC GCGCTATCGC CTCTACATCG CCGCGAAGA AACACAGTC
 751 TTGCTCATCG GCGCGACCCA AATCGAAGCG GAAGCCGAG CCGCGCGGAG
 25
 801 CGTACGTTCC GGGCTGGAAC TCTTATCCGC GCTCTATGCC GTCCACCGCG
 851 CTTTGGCGGA AGCGGACATC CTGGAATCG CGCGCGGCGT GCGCGCGGAG
 901 CTCACCCACC ACAACCCCGA AATCGCGTAC AGCGCGGACG GCGCGCTCAT
 951 CGAAATCAAC GGCCTTTTCC GGCACGCGTT TATGATTTCG CCGCGCGTAA
 1001 CGCGCGCGCG CGTCAGATTG GCAGTGGCAC TGTATTGCGG AAAAGACGCG
 30
 1051 CCGGACCGTG ATGAAGAAAG CGGTTTGGCG TATATCGGAA GACAAGATTA
 1101 A

This corresponds to the amino acid sequence <SEQ ID 818; ORF126ng-1>:

35
 1 MTRIAVLGGG LSGRLTAQL AEGQYQIELF DKGTRQGEHA AAYVAAAMLA
 51 PAASEAVEATP EVIRLGRQSI PLWRGIRCLR NLTMQENG SLIVWHGQDK
 101 PLSSEFVRHL KRGVADDEI VWRADSEIAE REPOLGGRFS DGIYLPTEGG
 151 LDGRQILSAL ADALDELNVF CHEHECAPQ DLQAQYDWVI DCRGYGAKTA
 201 WNQSPHSTST LRIGRGEVAR VYTPETILNR PVRLHPRYP LYAPKENHV
 251 FVIGATQIES EQAPASVRS GLELLSALYA VHPAFGEADI LEIAAGLRPT
 301 LNHNHPEIRY SRERRLIEIN GLFRHGFMS PAVTAAAVRL AVALFDGKDA
 40
 351 PERDEESGLA YIGRQD*

ORF126ng-1 and ORF126-1 show 95.1% identity in 366 aa overlap:

	10	20	30	40	50	60
orf126-1.pep	MTRIAVLGGGLSGRLTAQLAEQGYQIALFDKGRGRRGEHAAAYVAAAMLAPAEAVEATP					
orf126ng-1	MTRIAVLGGGLSGRLTAQLAEQGYQIELFDKGRGRRGEHAAAYVAAAMLAPAEAVEATP					
	10	20	30	40	50	60
orf126-1.pep	EVVRLGRQSI PLWRGIRCLRNLTMQENGSLIVWHGQDKPLSSEFVRHLKRGVADDEI					
orf126ng-1	EVIRLGRQSI PLWRGIRCLRNLTMQENGSLIVWHGQDKPLSSEFVRHLKRGVADDEI					
	70	80	90	100	110	120
orf126-1.pep	VWRADDDTAEREPLGGRFSDGIYLPTEGGDLGRQILSALADALDELNVFCHWEHECVPE					
orf126ng-1	VWRADDEIAEREPLGGRFSDGIYLPTEGGDLGRQILSALADALDELNVFCHWEHECAPQ					
	130	140	150	160	170	180
orf126-1.pep	GLQAQYDWVIDCRGYGAKTAWNQSPHSTSLRIGRGEVARVYTPETILNRPVRLHPRYP					
orf126ng-1	DLQAQYDWVIDCRGYGAKTAWNQSPHSTSLRIGRGEVARVYTPETILNRPVRLHPRYP					
	190	200	210	220	230	240

-450-

		250	260	270	280	290	300
	orf126-1.pep	LYIAPKENHVFVIGATQIESESQAPASVRSGLLELLSALYAHFAFGEADILEIATGLRPT					
5	orf126ng-1	LYIAPKENHVFVIGATQIESESQAPASVRSGLLELLSALYAHFAFGEADILEIATGLRPT					
		250	260	270	280	290	300
	orf126-1.pep	LNHHNFEIYRNRARRLIEINGLFRHGFMI SPAVTAARLVAALPDGKDAPEDEKESGLA					
10	orf126ng-1	LNHHNFEIYRSRRRLIEINGLFRHGFMI SPAVTAARLVAALPDGKDAPEDEKESGLA					
		310	320	330	340	350	360
15	orf126-1.pep	YIRRQDX					
	orf126ng-1	YIGRQDX					

Furthermore, ORF126ng-1 shows homology to a putative *Rhizobium* oxidase flavoprotein:

	gi 2627327 (AF004408) putative amino acid oxidase flavoprotein [Rhizobium etli]
20	Length = 327
	Score = 169 bits (423), Expect = 3e-41
	Identities = 112/329 (34%), Positives = 163/329 (49%), Gaps = 25/329 (7%)
	Query: 3 RI AVIGGGLSGRLTAIQLAEQGYQIELFDKGTQRQERXXXXXXXXXXXXXXXXXXXX 62
25	Sbjct: 2 RILVNGAGVAGLTVAWQLYRHGFRVTLAERAGTVGA-GASGFAGMLAPWCERESAEEP 60
	Query: 63 IRLGRQSIPLWRGIRCRINTLTMQENGSLIVWHGQDKPLSSEFVRHLKRGVADDEIVR 122
	+ LGR + W + G+L+V G-D F R G DE+
30	Sbjct: 61 LTLGRLAADWWEAA-----LPGHVHRRGTLVVAGGRDTGELDRFSRRTS-GWEWLDEVA- 113
	Query: 123 WRADEIAEREPEQLGGRFSDGIYLPTEGQLDGRQILSALADALDELNVPCWHEHECAPQDL 182
	IA EP L GRF ++ E LD RQ L+ALA L++ + +
	Sbjct: 114 -----IAALEPDLAGRFRRALFFRQEAHLDPQALAAAGLEDARMRLTLG---VVGVS 165
35	Query: 183 QAQYDWVIDCRGYGAKTAWNQSPHESTLRLGIRGEVARVYTPETLNNRPVLLHPRYPY 242
	+D V+DC G LRG+RGE+ V T E++L+RPVRLHPR+P+Y
	Sbjct: 166 DVDHDRVVDCTGAA-----QIGRLPGLRGVREMLCVETEVSLSRFPVLLHPRHPIY 218
40	Query: 243 IAPKENHVFVIGATQIESESQAPASVRSGLLELLSALYAVHFAFGEADILEIAAGLRPTIN 302
	I P++ + F++GAT IES- P + RS +ELL+A YA+HPAFGEA + E AG+RP
	Sbjct: 219 IVPRDKNRFMVGATMIESDDGGPITARSIMELLNAAAYAMHPAFGEARVETETGAGVRPAYP 278
	Query: 303 HHNFEIYRSRRRLIEINGLFRHGFMI SP 331
45	+ P R ++E R + +NGL+RHGF+++P
	Sbjct: 279 DNL P--RVTQEGRTLHVNGLYRHGFLLAP 305

This analysis suggests that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 97

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID

50 819>:

	1	ATGACTGATA	ATCGGGGGTT	TAOCTGGT	GAATTAATAT	CAGTGTCTTT
	51	GATATTGTCT	GTACTTGCTT	TAATTGTPTTA	TCOGAGCTAT	CGCAATTATG
	101	TTGAGAAAGC	AAAGATAAAT	GCAGTGGGGG	CAGCCTTTGT	AGAAAATGCA
	151	CATTTTATGG	AAAAGTTTTA	TCCTGCAGAT	GGGAGGTTTA	AACAAAACATC
55	201	TACCAAGTGG	CCAAGTTTGC	CGATTAAAGA	GGCAGAAGGC	TTTTGTATCC
	251	GTITGAATGG	AATCGTGGCG	CGGG .GCTT	TAGACAGTAA	ATTGATGTGT
	301	AAGCGGGTAG	CCATAGATAA	AGATAAAAAT	CCITTTATTA	TTAAGATGAA
	351	TGAAATCTTA	GTAACCTTTA	ATTTCGAGA	AGTCGSCAG	TTGCTGTAGT
	401	CACGGCTGG	ATTATTTTAA	AGGAATGAT	AAGGATCTCA	AGTTACTTTA
60	451	GTAG				

This corresponds to the amino acid sequence <SEQ ID 820; ORF127>:

```

1 MTDNRGFTLV ELISVVLLLS VLALIVYPSY RNYVEKAKIN AVRAALLENNA
51 HFMEKFYLN GRFKQTSTKW PSLPIKEAEG FCIRLNGIVA RXALDSKFML
101 KAVAIDKDN PFIKMNENL VTFICKKSAS CSDGLDYFK GNDKDKLLK
151 *

```

Further work revealed the following DNA sequence <SEQ ID 821>:

```

1 ATGACTGATA ATCGGGGGTT TACGCTGGT GAATTAATAT CAGTGGTCTT
51 GATATTGTCT GTACTTGCTT TAATTGTTTA TCCGAGCTAT CGCAATTATG
101 TTGAGAAAGC AAAGATAAAT CAGTCGCGG CAGCCTTGTT AGAAATGCA
151 CATTTTATGG AAAAGTTTTA TCTGCAGAA GGGAGGTTTA AACAAACATC
201 TACCAAGTGG CCAAGTTTGC CGATTAAAGA GGCAGAGGC TTTTGTATCC
251 GTTGAATGG AATCGCGCGC GGGGCTTTAG ACAGTAAATT CATGTTGAAG
301 GCGGTAGCCA TAGATAAAGA TAAAAATCT TTTATTATTA AGATGAATGA
351 AAATCTAGTA ACCTTTATTT GCAAGAGTC CGCCASTTCG TGTAGTGACG
401 GGCTGGATTA TTTTAAAGA AATGATAAG ACTGCAAGT ACTTAAGTAG

```

This corresponds to the amino acid sequence <SEQ ID 822; ORF127-1>:

```

1 MTDNRGFTLV ELISVVLLLS VLALIVYPSY RNYVEKAKIN AVRAALLENNA
51 HFMEKFYLN GRFKQTSTKW PSLPIKEAEG FCIRLNGIAR GALDSKFMLK
101 AVAIDKDN PFIKMNENL VTFICKKSAS CSDGLDYFK GNDKDKLLK*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF127 shows 98.0% identity over a 150aa overlap with an ORF (ORF127a) from strain A of *N.*

meningitidis:

```

25 orf127.pep MTDNRGFTLV ELISVVLLLSVLALIVYPSYRNYVEKAKINAVRAALLENNAHFMEKFYLN
orf127a MTDNRGFTLV ELISVVLLLSVLALIVYPSYRNYVEKAKINTVRAALLENNAHFMEKFYLN
30 orf127.pep GRFKQTSTKWPSLPIKEAEGFCIRLNGIVARXALDSKFMLKAVAIDKDNPFIIKMNENL
orf127a GRFKQTSTKWPSLPIKEAEGFCIRLNGI-ARGALDSKFMLKAVAIDKDNPFIIKMNENL
35 orf127.pep VTFICKKSASSCSDGLDYFKGNDKDKLLKX
orf127a VTFICKKSASSCSDGLDYFKGNDKDKLLKX
40 orf127a

```

The complete length ORF127a nucleotide sequence <SEQ ID 823> is:

```

1 ATGACTGATA ATCGGGGGTT TACGCTGGT GAATTAATAT CAGTGGTCTT
51 GATATTGTCT GTACTTGCTT TAATTGTTTA TCCGAGCTAT CGCAATTATG
101 TTGAGAAAGC AAAGATAAAT CAGTCGCGG CAGCCTTGTT AGAAATGCA
151 CATTTTATGG AAAAGTTTTA TCTGCAGAA GGGAGGTTTA AACAAACATC
201 TACCAAGTGG CCAAGTTTGC CGATTAAAGA GGCAGAGGC TTTTGTATCC
251 GTTGAATGG AATCGCGCGC GGGGCTTTAG ACAGTAAATT CATGTTGAAG
301 GCGGTAGCCA TAGATAAAGA TAAAAATCT TTTATTATTA AGATGAATGA
351 AAATCTAGTA ACCTTTATTT GCAAGAGTC CGCCASTTCG TGTAGTGACG
401 GGCTGGATTA TTTTAAAGA AATGATAAG ACTGCAAGT ACTTAAGTAG

```

This encodes a protein having amino acid sequence <SEQ ID 824>:

```

1 MTDNRGFTLV ELISVVLLLS VLALIVYPSY RNYVEKAKIN TVRAALLENNA
51 HFMEKFYLN GRFKQTSTKW PSLPIKEAEG FCIRLNGIAR GALDSKFMLK
101 AVAIDKDN PFIKMNENL VTFICKKSAS CSDGLDYFK GNDKDKLLK*

```

ORF127a and ORF127-1 show 99.3% identity in 149 aa overlap:

		10	20	30	40	50	60
5	orf127a.pep	MTDNRGFTLV	ELISVVLIL	SVLALIV	PSYRNYVEK	AKINAVRAAL	LENAHFMEK
	orf127-1	MTDNRGFTLV	ELISVVLIL	SVLALIV	PSYRNYVEK	AKINAVRAAL	LENAHFMEK
		10	20	30	40	50	60
		70	80	90	100	110	120
10	orf127a.pep	GRFKQTSTK	WPSLPIKEA	EGFCIRL	NGIARGALD	SKFMLKAVA	IDKDNPF
	orf127-1	GRFKQTSTK	WPSLPIKEA	EGFCIRL	NGIARGALD	SKFMLKAVA	IDKDNPF
		70	80	90	100	110	120
		130	140	150			
15	orf127a.pep	TFICKKSASS	CSGDLDF	FKGNDK	CKLLKX		
	orf127-1	TFICKKSASS	CSGDLDF	FKGNDK	CKLLKX		
		130	140	150			

20 Homology with a predicted ORF from *N.gonorrhoeae*

ORF127 shows 97.3% identity over a 150 aa overlap with a predicted ORF (ORF127ng) from *N.gonorrhoeae*:

	orf127.pep	MTDNRGFTLV	ELISVVLIL	SVLALIV	PSYRNYVEK	AKINAVRAAL	LENAHFMEK	FYLQN	60
25	orf127ng	MTDNRGFTLV	ELISVVLIL	SVLALIV	PSYRNYVEK	AKINAVRAAL	LENAHFMEK	FYLQN	60
	orf127.pep	GRFKQTSTK	WPSLPIKEA	EGFCIRL	NGIARGALD	SKFMLKAVA	IDKDNPF	FIKMNENL	120
	orf127ng	GRFKQTSTK	WPSLPIKEA	EGFCIRL	NGI-ARGALD	SKFMLKAVA	IDKDNPF	FIKMNENL	119
30	orf127.pep	VTFICKKSASS	CSGDLDF	FKGNDK	CKLLK				150
	orf127ng	VTFICKKSASS	CSGDLDF	FKGNDK	CKLLK				149

The complete length ORF127ng nucleotide sequence <SEQ ID 825> is:

35	1	ATGACTGATA	ATCGGGGGTT	TACACTGGTT	GAATTAATAT	CAGTGGTCTT	
	51	GATATTGTCT	GTACTTGCTT	TAATTGTTTA	TCCGAGCTAT	CGCAATTATG	
	101	TTGAGAAAGC	AAAGATAAAT	GCAGTGGCGG	CAGCGTCTTT	AGAAATGCA	
	151	CATTTTATG	AAAGTTTGA	TCTGCAGAA	TGGAGATTTA	AACAAACATC	
	201	TACCAAATG	CCAAGTTTGC	CGATTAAAGA	GGCAGAAGGC	TTTGTATCC	
40	251	GTTTGAATG	AAATCGGCGC	GGGGCTTTAG	ACAGATAAAT	CATGTTGAAG	
	301	GGCGTAGCCA	TAGATAAAGA	TAAAAATCTC	TTTATTATTA	AGATGAATGA	
	351	AAATCTAGTA	ACCTTTAATT	GCAAGAAGTC	CGCCAGTTGC	TGTAGTGACG	
	401	GGCTGGATTA	TTTTAAAGA	AATGATAAGG	ACTGCAAGTT	ACTTAAGTAG	

This encodes a protein having amino acid sequence <SEQ ID 826>:

45	1	MTDNRGFTLV	ELISVVLIL	SVLALIV	PSYRNYVEK	AKINAVRAAL	LENAHFMEK	FYLQN
	51	HFMEK	FYLQN	GRFKQTSTK	WPSLPIKEA	EGFCIRL	NGIARGALD	SKFMLK
	101	AVAI	DKDNPF	FIKMNENL	TFICKKSASS	CSGDLDF	FKGNDK	CKLLKX*

ORF127ng and ORF127-1 show 100.0% identity in 149 aa overlap:

50	orf127-1.pep	MTDNRGFTLV	ELISVVLIL	SVLALIV	PSYRNYVEK	AKINAVRAAL	LENAHFMEK	FYLQN
	orf127ng-1	MTDNRGFTLV	ELISVVLIL	SVLALIV	PSYRNYVEK	AKINAVRAAL	LENAHFMEK	FYLQN
		10	20	30	40	50	60	
		70	80	90	100	110	120	
55	orf127-1.pep	GRFKQTSTK	WPSLPIKEA	EGFCIRL	NGIARGALD	SKFMLKAVA	IDKDNPF	FIKMNENL
	orf127ng-1	GRFKQTSTK	WPSLPIKEA	EGFCIRL	NGIARGALD	SKFMLKAVA	IDKDNPF	FIKMNENL
		70	80	90	100	110	120	

-453-

5
 10
 15
 20
 25
 30
 35
 40
 45
 50
 55
 60
 65
 70

```

    orf127-1.pep    TFICKKSASSCS DGLDYFKGNDKDKLLKX
                   |||
    orf127ng-1     TFICKKSASSCS DGLDYFKGNDKDKLLKX
                   |||
  
```

This analysis, including the fact that the predicted transmembrane domain is shared by the meningococcal and gonococcal proteins, suggests that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

10 Example 98

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 827>

1
 5
 10
 15
 20
 25
 30
 35
 40
 45
 50
 55
 60
 65
 70

```

    1  ..GTGTCGCTGG CTTGGGTGAT TGCCTCTCAA ATCTTCCTTT ACGAAGATT
    5  CAACCAATG CGGAAACCG GTGGAGCTAT CTGCGGTTTT CTTGTCCAA
    10 ATTTATCTGG GGTTCACAG GGGGTATTT GATTTGAGTG CCGACGAGAA
    15 CCGCGTACTG CATATCTGTT CTTTGGCAGT AGAGGAACAG TATTACCTCC
    20 TGTATCCCTT TTTGCTGATA TTTTGTGCTA AAAAACCAG ATCGCTACGG
    25 GTGCTGGCTA ACATCAGCAT CATCTGCTTT TTGATTTTGA CTGCTCATCT
    30 GTTTTGGCCA AGCGGGTCTT ATACCGACAT CCTCAACCAA CCCAATCAT
    35 ATTACCTTTC GACACTGAGG TTTCCGAGCG TGTGGCAGG TTCGCTGCTG
    40 GCGGTTTACG GGCACACGCA AAACGCGCAG CCGCACAACAG CAAATGGAAA
    45 ACGGCGATTG CTTTATCATC TGTGCTTCGG CGCATTTGCTT GCTGCTGCTT
    50 TCGTGTATTGA CAACACAAAT CCGTTTATCC CGGGAATGAC CCGTCTCTCT
    55 CCGTGGCTGC TGAACGCACT GCTTATCCGG AGTATGCAAT ACGGACACT
    60 TCCGACCCGC ATCTCTGCGG CAAGCCCATC CGTATTGTCT GCACAAATCT
    65 CTTATTCCTC ATACCTGTAC CATTTGATTT TTATTGCTTT CGCTCGCGTC
    70 ATTAGAGGCG GGAACACAGT CGGACTGCCT GCCG..
  
```

This corresponds to the amino acid sequence <SEQ ID 828; ORF128>:

1
 5
 10
 15
 20

```

    1  ..VSLASVIA SQ IFLYEDFNQM RKTVELSAVF LSNLYLGFQQ GYFDLSADEN
    5  PVLHIWSLAV EQYQYLLYPL LLIFCCCKTK SLRVLRNISI ILFLILTASS
    10 FLPSGFYTDI LNQENTYYLS TLRFPPELLAG SLLAVYGGQT NGRROTANGK
    15 RQLLSLCLFG ALLACLFLVD KHNPFIPGMT LLLPCLLTAL LIRSMQYGTL
    20 PTRILSASPI VFVGKISYSL YLHWIFIAF APLIRGGKQL GLPA..
  
```

Further work revealed the complete nucleotide sequence <SEQ ID 829>:

1
 5
 10
 15
 20
 25
 30
 35
 40
 45
 50
 55
 60
 65
 70
 75
 80
 85
 90
 95
 100
 105
 110
 115

```

    1  ATGCAAGCTG TCCGATACAG ACGGAAATTT GACGGATTGC GGGCCGTCGC
    5  CGTGCTATCC GTCATGATTT TCCACCTGAA TAACCGCTGG CTGCCCGGAG
    10 GATTCTCTGG GGTGGACATT TCTTTTGTCA TCTCAGGATT CACTCATTAAC
    15 GGCATCATTC TTCTGAAAT ACAGAACGCT TCTTTTCTT TCCTGGATT
    20 TTATACCCGC AGGATTAAGC GGAATTAATC TGCTTTATT CGCGCCGTGT
    25 CGCTGGCTTC GGTGATTGCG TCTCAAAATCT TCCTTACGTA AGATTTCAAC
    30 CAAATGCGGA AAACCGTGGG GCTTTCTGCG GTTTTCTGT CCAATATTTA
    35 TCTGGGGTTT CAGCAGGGGT ATTTCTGATT GAGTGCCGAC GAGAACCCCG
    40 TACTGCATAT CTGCTCTTTG CAGATGAGGG AACAGTATTA CTTCTCTGAT
    45 CCCCTTTTGC TGAATTTTTC CTGCAAAAAA ACCAAATGCG TACGGGTGCT
    50 GCGTAACATC AGCATCATCC TGTTTTGAT TTTGACTCGC TCATCTGTTT
    55 TGCAAGCGGG GTTTTATACC GACATCCTCA ACCAACCCAA TACTTATTAC
    60 CTTTGACAC TGAAGTTTTC CGAGCTGTTG CGAGGTTCGC TGCTGGGGGT
    65 TTACGGSCAA ACGCAAAACG CGAGACGGCA AACAGCAAA TGGAAAACGGC
    70 AGTGCTCTTC ATCACTCTGC TCGCGCGCAT TGCTTGCCTG CCGTGTGTGT
    75 ATTGACAAAC ACAATCCGTT TATCCCGGSA ATGACCCCTG TCTCTCCCTG
    80 CTTGCTGAGC GTCGCTGTA TCCGCGATAT GCAATAGGG ACATTCGGA
    85 CCCGCACTCT GTGCGCAAGC CCCATCGTAT TTGTCGCAAA AATCTCTTAT
    90 TCCCTATACC TGTACCATGT GATTTTATT GCTTTGCCCC ATTACATTAC
    95 AGGCGACAAA CAGCTGGGAC TGCTTCGGGT ATCGGCGGTT GCGCGCTTGA
    100 CGGCGCGATT TTCCCTGTTG AGTTATTATT TGATTGAACA CGCGCTTAGA
    105 AAACGGAAGA TGACCTTCAA AAGGCGATT TTCTGCTCTT ATCTCGCCCC
    110 GTCCCTGATA CTTTGCSTTT ACACCTGTA CGCAAGGGGG ATATTGAAC
    115 AGGAACACT CCGCCCGTTG CCGGCGGCGC CCCTTGCCTG GGAATAATCAT
  
```

-454-

1201 TTTCGGGAAA CCGTCCTGAC CCTCGGCGAC TCGCAGCGCG GACACCTGAG
 1251 GGGGTTTCTG GATTATGTCG GCAGCCGGGA AGGGTGGAAA GCCAAAATCC
 1301 TGTCCTCGA TCGGAGTGT TTGGTTGGG TAGATGAGAA GCTGGCAGAC
 1351 AACCCSTTAT GTCGAAAATA CCGGGATGAA GTTGAAAAAG CCGAAGCCGT
 1401 TTTCATTGCC CAATCTCATG ATTTGAGGAT GGGCGGCCAG CCTGTGCCAA
 1451 GATTTGAGGC SCATCTCTTC CTATATCCCG GGTTCGCCAG CCGATTCCAG
 1501 GAACCGCTCA AAGCATGAGC CGCCTGAA CCGCTCATG TTTTTCGAAA
 1551 CACACATCA ATCAGCGGTT CGCCCTGAG GAGAGAAAAA TTGAAAAGAT
 1601 TTGCGGCAAA CCAATATCTC CGCCCATTC AGGCTATGG CGACATCGG
 1651 AAGAGCAATC AGCGGCTCTT TGATTTGATT AAAGATATTC CCAATGTGCA
 1701 TTGGGTGGAC GCACAAAAAT ACCTGCCCAA AAACACGGTC GAAATATACG
 1751 GCGCGTATCT TTACGGCGAC CAAGACCACC TGACCTATT CGSTTCTTAT
 1801 TATATGGGCG GGAATATCCA CAAACACGAA CGCCTGCTTA AATCTTCCCA
 1851 CGCGCGCGCA TTGCAGTAG

15 This corresponds to the amino acid sequence <SEQ ID 830; ORF128-1>:

1 MQAVRYRPEI DGLRAVAVL VMIFHLNNRW LPGSFLGVDI FFVISGLFIT
 51 GIILSEIQNG SFSFRDFYTR RIKRIYPAPI AAVSLASVIA SQIFLYEDFN
 101 QMRKTVELSA VLSNIYLFQ QGYFDLSAD ENPVLIHWSI AVEEQYLLY
 151 PLLILFCCKK TKSRLVLRNI SIILFLILTA SSFLPSGFYT DIINQPNITY
 201 LSTLRPELL AGSILAVYGQ TQNGRRQTAN GRQILSSLC FGALLACLIV
 251 IDKNPPIFG MTLLPCLLT ALLRSQMVG TLPTRILAS PTUVFGKTSY
 301 SLIYLHWIFI AFAYITGDK QLGLPAVSAA AATAGTFSLL SYVLEQPIRL
 351 KRKMTFKKAF FCYLLAPSLI LVGNLYARG ILKQEHRLRP PGAPLAENH
 401 FPEVTILTGD SHAGHLRGL DYVSGREGWK AKILSLDSEC LVWVDEKLAD
 451 NPLCRKYRDE VEKAEAVFIA QFYDLRMGGQ PVPRFEAGS LIPGFARFR
 501 ETVKRIAAVK PVYVFNANTS ISRSPLREEK LKRFAANQYL RPIQAMGDIG
 551 KSNQAVFDLI KDIPNVHWD AQYLPKNTV EIYGRYLYGD QDHLTYFGSY
 601 YMGREFHKHE RLLKSSHGGA LQ*

Computer analysis of this amino acid sequence gave the following results:

30 Homology with hypothetical integral membrane protein HI0392 of *H. influenzae* (accession number U32723)
 ORF128 and HI0392 show 52% aa identity in 180aa overlap:

Orf128: 1 VSLASVIASQIFLYEDFNQMRKTVELSAVLSNIYLFQGGYFDLSADENPVLIHWSLAV 60
 +L S IAS IF+Y DFN++RKT+EL+ FLN YLG QGYFDLSA+ENPVLIHWSLAV
 35 HI0392: 46 MALVSFIASAIFFIYDNFNKRLKRTIELAIFLSNFIYLGTLQGYFDLSANENPVLIHWSLAV 105
 Orf128: 61 EEQXXXXXXXXXIFCCKKTKSLRVLNISIILFLLTASSFLPSGFYTDILNQPNITYLS 120
 E Q I KK + +VL I++ILF IL A+SE+ + FY +L+QPN YYLS
 HI0392: 106 EQQYLIIFLLILILAYKKFREVKVLFITLILFILLTASSFLPSGFYTDILNQPNITYLS 165
 Orf128: 121 TLRPELLAGSLLAVYGQTQNGRRQTANGKRLQLSSLCFGALLACLEVDIKHNPFIPIGMT 180
 LRFPELL GSLLA+Y N + Q + +L+ L L +CLF+++ + FIPG+T
 HI0392: 166 NLRPELLVGSLLAIYHNLSN-KVQLSKQVNNILAILSTLLFSCLFIMNNNIAFIPGIT 224

Homology with a predicted ORF from *N. meningitidis* (strain A)

45 ORF128 shows 98.0% identity over a 244aa overlap with an ORF (ORF128a) from strain A of *N. meningitidis*:

orf128 pep 10 20 30
 VSLASVIASQIFLYEDFNQMRKTVELSAV
 50 orf128a ILSEIQNGSFSFRDFYTRRIKRIYPAPIAAVSLASVIASQIFLYEDFNQMRKTVELSAV
 60 70 80 90 100 110
 40 50 60 70 80 90
 orf128 pep LSNIIYLFQGGYFDLSADENPVLIHWSLAVEEQYLLYPLLIFCCKKTKSLRVLNISI
 55 orf128a LSNIIYLFQGGYFDLSADENPVLIHWSLAVEEQYLLYPLLIFCCKKTKSLRVLNISI
 120 130 140 150 160 170
 100 110 120 130 140 150
 60 orf128 pep ILFLILTASSFLPSGFYTDILNQPNITYLSTLRPELLAGSLLAVYGQTQNGRRQTANGK
 110 120 130 140 150

orf128a	ILFLILITATSLFSPSGFYTDILNQPNITYLSTLRFPELLAGSLLAVYQGTQNGRRQTANGK	180	190	200	210	220	230
5	orf128.pep	RQLSSSLCFGALLACL FVIDKHNFPIGMTLLLPCLLTALLIRSMQYGTLPTRILSASPI	160	170	180	190	200
	orf128a	RQLSSSLCFGALLACL FVIDKHNFPIGMTLLLPCLLTALLIRSMQYGTLPTRILSASPI	240	250	260	270	280
10	orf128.pep	VFVGKISYSLYLHWIFIAFAELIRGGKQLGLPA	220	230	240		
	orf128a	VFVGKISYSLYLHWIFIAFAHYITGDKQLGLPAVSVAALTAGFSLLSYLIEQPLRKR	300	310	320	330	340
15	orf128a	KMTFKKAFCLYLAPSLILVGYNLYARGILKQZHLRPLGAPLAAENHFPETVLTLDGDSH	360	370	380	390	400
						410	

The complete length ORF128a nucleotide sequence <SEQ ID 831> is:

20	1	ATGCAAGCTG	TCGCATACAG	ACCGGAAATT	GACGATTGCG	GGCGCGTCCG
	51	CGTGCATAC	CTCAGCATTT	TCCACCTGAA	TACCGCGTGG	CTGCGCGGAG
	101	GATTCCTGGG	GGTGACATTT	TCTTTTGTCA	TCCAGGATTT	CCTCATTACC
	151	GGCATCATT	TTTCTGAAAT	ACAGAACGGT	TCTTTTCTTT	TCCGGGATTT
	201	TTATACCCCG	AGGATTAAAG	GGATTATACC	TGCTTTTAT	GCGGCGGTGT
	251	CGCTGGCTTC	GGTGAATTGC	TCTCAATCT	TCTTTCTAG	AGATTTCAC
25	301	CAAAATCGGA	AAACCGTGGG	GCTTTCTGCG	GTTTCTGTGT	CCAATATTAT
	351	TCTGGGGTTT	CAGCAGGGGT	ATTTCGATTT	GAGTGCCTAG	GAGAACCCCG
	401	TACTGCATAT	CTGGTCTTTG	GCAGTAGAGG	AACAGTATTA	CCTCGTGTAT
	451	CCTCTTTTGC	TGATATTTTG	CTGCAAAAAA	ACAAAATCGC	TACGGGTGCT
	501	GGGTAAACAT	AGCATCATCC	TATTTCTGAT	TTGACTGCG	ACATCGTTTT
30	551	TGCCAAGGGG	GTTTTATACC	GATATCTCTCA	ACCAACCCAA	TACTATTATC
	601	CTTTCGACAC	TGAGGTTTCC	CGAGCTGTTG	CGAGGTTGCG	TGCTGGCGGT
	651	TTACGGGCAA	ACCGAARACG	CGACAGCGCA	AAACAGCAAT	GGAAAACGCG
	701	AGTGGTCTTC	CTACCTCTGC	TTCCGCGCAT	TGCTTTCTGT	CCGTTGCTGT
	751	ATTCGACAC	ACAATCCGTT	TTTCCCGGGA	GAACCTCTGT	TCTTCCCTGT
35	801	CCTGCTGAAG	GCACCTGCTTA	TCCGCGATAT	GCATACCGG	ACACTTCCGA
	851	CCCGCATCCT	GTCGGCAAGC	CCCATGATAT	TTGTCGGCAA	AATCTCTTAT
	901	TCCCTATACC	TGTACCAATTG	GATTTTATTT	GCTTTCGCCG	ATTACATTAC
	951	AGGCGACAAA	CAGCTCGGAC	TGCTCGCGGT	ATCGGCGGTT	GCGCGTTAGA
40	1001	CGGCGGGAAT	TTCCCTGTTG	AGTTTATTTT	TGATTGAACA	GCGCGTTAGA
	1051	AAACGGGAAGA	TGACCTTCAA	AAAGGCATTT	TTTCTGCTCT	ATCTCGCCCC
	1101	GTCCCTGATA	CTTGTCCGTT	ACAACCTGTA	CGCAAGGGGG	ATATTGAAC
	1151	AGGAACACCT	CGGCCCGTTG	CCGCGCGGCG	CCCTTGTCTG	GGAAAATCAT
	1201	TTTCGGGAAA	CGGCTCTGAC	CTCGGCGGAC	TGCGACGGCG	GACACCTGCG
	1251	GGGGTTTCTG	GATTATGTGC	GCAGCGGGGA	AGGGTGGAAA	GCCAAATATC
45	1301	TGTCCTTCA	TTCCGAGTGT	TGCTGTTTGG	TAGATGAGAA	CGTGGCGAG
	1351	AACCCGCTAT	GTGCAARATA	CCGGATGAGG	GTTGAAAGAG	CCGAGCCGCT
	1401	TTTCATTGCC	CAATCTCATG	ATTTCAGAT	GGGCGGCGAG	CCGCTGCCGA
	1451	GATTTGAAGC	GCAATCTCTC	CTAATACCGG	GGTTCGCCAG	CCGATTTCAG
	1501	GAAACCGTCA	AAAGCATAGC	CGCCCTCACA	CCGCTCTATG	TTTTTGCAAA
50	1551	CAACACATCA	ATCAGCCGTT	CGCCCTGAG	GGAGGAAAAA	TTGAAAGAT
	1601	TTGCGCGAAA	CCAATATCTC	CGCCCATTC	AGGCTATGGG	GCACATCGCG
	1651	AAGAGCAATC	AGGCGGCTCT	TGATTTGATT	AAAGATATTC	CCAATGTGCA
	1701	TTGGGTGGAC	GCACAAAAAT	ACCTGCCCAA	AAACACGGTC	GAAATATACG
	1751	GCGGCTATCT	TTACGGCGAC	CAAGACACCC	TGACCTATTT	CGGTTCTTAT
55	1801	TATATGGGGG	GGGAATTTCA	CAAAACAGAA	CGCCTGCTTA	ATCTTCTCTG
	1851	CGAAGCGGCA	TTGCAAGTAG			

This encodes a protein having amino acid sequence <SEQ ID 832>:

60	1	MQAVRYRPEI	DLCLRAVAVL	VMIFHIANRW	LPGGFLGVDI	FPVISGFLIT
	51	GIILSELQNG	SFSRDFYTR	RIKRYIPAPL	AAVSLASVIA	SDILYEDFN
	101	QMRKTVLSLA	VFLSNLYLFG	QQCYFDLSAD	ENFVLHWSL	AVEEQYVLY
	151	PIILLIFCCCK	TKSLRLVLNI	SIILFILITA	TSFLPSGPHY	DILNQPNITY
	201	LSTLRFPELL	AGSLAVLYQG	TQNGRRQTAN	KRQRLSSLC	FGALLACLFL
	251	IDKHNFPIPG	MTLLLPCLLT	ALLIRSMQYV	TLPRILSAS	PIVVFVKISY
	301	SLYLYHWIET	AFARHYITGDK	QLGLPAVSVA	AALTAGFSLL	SYLYIEQPLR
65	351	KRKMTFKKAF	FCLYLAPSLI	LVGYNLYARG	ILKQZHLRPL	PGAPLAAENH
	401	FPETVLTLDG	SHAGHLRGLF	DYVGSREGWK	AKILSLDSEC	LWVWDKLAD
	451	NPLCRKYRDE	VEKAEAVFIA	QFYDLRMGGQ	PVPRFEAQSF	LIPGFPARFR

```

501  ETVKRIAALK PVYVFANNTS ISRSPLEEK LKRFANQYL RPIQAMGDIG
551  KSNQAVFDLI KDIPNVHWD ACKYLPKNTV EYGRYLYGD QDHLTYFGSY
601  YMGREFHKE RLLKSSRDGA LQ*

```

ORF128a and ORF128-1 show 99.5% identity in 622 aa overlap:

```

5  orf128a.pep  MQAVRYRPEIDGLRAVAVLSVMIFHLNNRWLPGGFLGVDIFFVISGFLITGIILSEIQNG
    orf128-1    MQAVRYRPEIDGLRAVAVLSVMIFHLNNRWLPGGFLGVDIFFVISGFLITGIILSEIQNG

10 orf128a.pep  SFSFRDFTYTRIKRIYPAFIAAVSLASVIASQIFLYEDFNQMRKTVELSAVLSNIYLG
    orf128-1    SFSFRDFTYTRIKRIYPAFIAAVSLASVIASQIFLYEDFNQMRKTVELSAVLSNIYLG

15 orf128a.pep  QQGYFDLSADENPVLHIWLSLAVEEQYLLYPILLIFCCKTKSLRVLRNISIIFLILTA
    orf128-1    QQGYFDLSADENPVLHIWLSLAVEEQYLLYPILLIFCCKTKSLRVLRNISIIFLILTA

20 orf128a.pep  TSFLPSGFTYDIINQPNYYLSTLRFPPELLAGSLAVYGGTQNGRRQTANGKRQLLSSLC
    orf128-1    TSFLPSGFTYDIINQPNYYLSTLRFPPELLAGSLAVYGGTQNGRRQTANGKRQLLSSLC

25 orf128a.pep  FGALLACLFVIDKHNPFIPGMTLLLPCLLTALLIRSMQYGTLPTRILSASPIVFGKISY
    orf128-1    FGALLACLFVIDKHNPFIPGMTLLLPCLLTALLIRSMQYGTLPTRILSASPIVFGKISY

30 orf128a.pep  SLYLHNIFIAFAHYITGDKQLGLPAVSAVAALTAGFSLSYLIEQPLAKRMTFKKAF
    orf128-1    SLYLHNIFIAFAHYITGDKQLGLPAVSAVAALTAGFSLSYLIEQPLAKRMTFKKAF

35 orf128a.pep  FCYLYAPSLILVGNLYARGILKQEHRLPLGAPLAENHFPETVLTIGDSHAGHLRGFL
    orf128-1    FCYLYAPSLILVGNLYARGILKQEHRLPLGAPLAENHFPETVLTIGDSHAGHLRGFL

40 orf128a.pep  DYVSGREGWKAKILSLDSECLVWVDEKLADNPLCRKRYRDEVEKAEAVFIAQFYDLRMGGQ
    orf128-1    DYVSGREGWKAKILSLDSECLVWVDEKLADNPLCRKRYRDEVEKAEAVFIAQFYDLRMGGQ

45 orf128a.pep  PVPRFEAQSFILPGFPAFRFETVKRIAALKPVYVFANNTSISRSPLREKLRFAANQYL
    orf128-1    PVPRFEAQSFILPGFPAFRFETVKRIAALKPVYVFANNTSISRSPLREKLRFAANQYL

50 orf128a.pep  RPIQAMGDIGKSNQAVFDLIKDI PNHVHWDQAKYLPKNTVEIYGRYLYGDQDHLTYFGSY
    orf128-1    RPIQAMGDIGKSNQAVFDLIKDI PNHVHWDQAKYLPKNTVEIYGRYLYGDQDHLTYFGSY

55 orf128a.pep  YMGREFHKEHLKSSRDGALQX
    orf128-1    YMGREFHKEHLKSSSHGALQX

```

Homology with a predicted ORF from *N.gonorrhoeae*

50 ORF128 shows 93.4% identity over 244 aa overlap with a predicted ORF (ORF128ng) from *N. gonorrhoeae*:

```

    orf128.pep  VSLASVIASQIFLYEDFNQMRKTVELSAVF 30
    orf128ng    ILSEIQNGSFSFRDFTYTRIKRIYPAFIAAVSLASVIASQIFLYEDFNQMRKTIELSTVF 112

55 orf128.pep  LSNIVLGQQGYFDLSADENPVLHIWLSLAVEEQYLLYPILLIFCCKTKSLRVLRNISI 90
    orf128ng    LSNIVLGFRGLGYFDLSADENPVLHIWLSLAVEEQYLLYPILLIFCYKTKSLRVLRNISI 172

60 orf128.pep  ILFLILTASSFLPSGFTYDIINQPNYYLSTLRFPPELLAGSLAVYGGTQNGRRQTANGK 150
    orf128ng    ILFLILTASSFLPAGFTYDIINQPNYYLSTLRFPPELLVGLAVYGGTQNGRRQTENGK 232

65 orf128.pep  ROLLSSLCFGALLACLFVIDKHNPFIPGMTLLLPCLLTALLIRSMQYGTLPTRILSASPI 210
    orf128ng    ROLLSSLCFGALLVCLFVIDKHDPFIPGMTLLLPCLLTALLIRSMQYGTLPTRILSASPI 292

```

```

orf128.pep  VFVGKISYSLYLHWIFAFAPLIRGGKQLGLPA      244
|||||
orf128ng    VFVGKISYSLYLHWIFAFAYHTGDKQLGLPAVSAVAALTAGFSLLSYLLIEQLPKRR 352

```

5 The complete length ORF128ng nucleotide sequence <SEQ ID 833> is:

```

1  ATGCAAGCTG TCCGATACAG GCCTGAATTT GACGGATTGC GGGCGGTGCG
51  CCGTGCATATC GTACATATTT TCCACCTGAA TACCGCTGCG CTGCGCGGAG
101  GATTCCTGGG GGTGACATATT TTCTTTTGCA TCCCGGGATT COTCATTTAC
151  AACATCATTC TTCTCGAAT ACAGAACGGT TCTTTTCTTT TCCGGGATTT
10  201  TTATACCCCG AGGATTAAAG GGAATTTATC TGCTTTTATT CGGCGCGTGT
251  CCGTGGCTTC GGTGATTGCT TCTCAATCTC TCGTTTACGA AGATTTCACAC
301  CAAATGAGGA AAACCATAGA GCTTTCTACG GTTTTTTGT CCAATATTTA
351  TTTGGGGTTC CGATTGGGGT ATTTGATTTT GAGTGCCGAC GAGAACCCTCG
401  TACTGCATAT CTGGTCTTTG GGGGTAGAGG AACAGTATTA CCTCTGTAT
15  451  CCTCTTTTGC TGATATTTCTG TTACAAAAAA ACCAATCAC TACGGGTGCT
501  GCGTAATATC AGCATCATCC TGTTTCTGAT TTTGACCGCA TCATCGTTTT
551  TGCGCGGCGG GTTTTATACC GACATCTCTA ACCAACCCaa TACTTATTAC
601  CTTTGACAC TGGAGTTTCC CGAGCTGTG GTGGGTTGCG GTTGGCGGTG
651  TTACGGGCAA ACAGAAAAGC CGAGACGGCA AACAGAAAAT GGAAACCGGC
701  AGTTCCTATC ATTACTCTGT TCGGCGCAT TCGTTGCTG CCGTGTCTGTG
751  ATCGACAAAG ACAGATCGTT TATCCGGGGA ATACACCGTC TCTTCCCTG
801  CCGTCTGACG CGCTGCTCTA TCCGGAGTAT GCAATACGGG ACATCTCCGA
851  CCGCGATCCT GTGCGCAAGC CCGATCGTAT TTGTCGGCAA AATCTCTTAT
901  TCCCTATACC TGTACCATTT GATTTTATT GCCTTCGCGC ATTACATTAC
951  AGGCGACAAA CAGCTCGGAG TGCTGCGGT ATCGCGGTTG CGCGCTTGA
1001  CGCGCGGATT TTCCCTGTTG AGCTATTATT TGATTGAACA GCCGCTTAGA
1051  AAACCGGAAGA TGACCTTCAA AAAGGCATTT TTCTGCGTT ATCTCGCCCC
1101  GTCCCTGATG CTTGTCGGTT ACAACCTGTA TTCAAGAGGG ATATTGAAAC
1151  AGGAACACCT CGCGCCGCTG CCGGCAACGC CCGTTGCTGC GGAAATAAAT
1201  TTTCCGGAAA CCGTCTTGAC CCGCGGCGAC TCGCAACGGG GACACCTGGG
1251  GGGGTTTCTG GATTATGTCG GCGGCAAGGA AGGGTGGAAA GCTAAAATCC
1301  TGTCCTCTGA TCCGAGATGT TGGGTTGGG TGGATGAGAA CTTGGCAGAC
1351  AACCAAAATA CCGGAGGAGA GTTGAAGGAA CCGAGAGCTG CCGAGATCTG
1401  TTTCATTGCG CAATTCTATG ATTTGAGGAT GGGCGGCGAG CCGTGGCGGA
35  1451  GATTTTGAAGC GCAATCCTTC CTGATACCGG GGTTCAAAGC CCGATTCAAG
1501  GAAACCGTGA AGAGGATAGC CGCGCTCAAA CCGTGATATG TTTTTCGAAA
1551  CAATACATCA ATCAGCGGTT CTCCCTTAGG GGAGGAAAAA TTGAAAAGAT
1601  TTGCTATAAA CCAATACCTC CGGCTATTTC GGGCTATGG CGACATCGGC
1651  AAGAGCAATC AGGCGGTCTT TGATTTGGTT AAAGATATTC CCAATGTGCA
40  1701  TTGGGTGGAC GCACAAAAAT ACCTGCCCAA AAACAAGGCT GAAATACACG
1751  GACGCTATCT TTAACGGGAC CAAGACCAAC TGACCTATTG CCGTCTCTAT
1801  TATATGGGGG GGAATTTTCA CAAACAGAA CGCCTGCTCA AGCATTCCCG
1851  AGGCGGCGCA TTGCAAGTAG

```

This encodes a protein having amino acid sequence <SEQ ID 834>:

```

45  1  MQAVRYRPEI DGLRAVAVLV VILFIANNRW LPGGFLGVDI PFVISGFLIT
51  NIILSEIQNG SFSFRDFTTR RIKRYPATET AAVGLASVIA SDIPLVEDFN
101  QMRKTIELST VFLSNLYLPG PLGYDLSDAD ENPVLHWLSL AVERGYLLY
151  FLLLIICYKK TKSRLVLRNI SIILFILTA SFLPAGPFTY DILNQNTYY
201  LSTLRPEELL VGSLLAVYQG TQNGRROTEN GRROLLSLLC PGALLVCLFV
50  251  IDKHDPFITP ITLLLECLLT ALLIRSMQVG TLPTRILSAS PIVFVGKISY
301  SLYLYHWIFI AFAYHTITGDK QLGLPAVSAV AALTAGFSLL SYLLIEQLPR
351  KRKMTFKKAF FCYLYLAPSLM LVGNYLSRSG ILKQELHRLP PGTPTVAENN
401  FPEVTVLGSD SHAGHLRGFL DYVGGREGWK AKILSDSECE LVWVDEKLD
55  451  NFLCRKYRDE VEKAEAVPIA QFYDLRMGGQ PVPRFAEQSF LIPGFKARF
501  ETVKRIAAVK PVVYFANNST ISRSLREEK LKRFAINQYL RPIRAMGDIG
551  KSNQAVFDLV KDIPNVHVDV AQRYLPKNTV EIHGRVLYGD QDHLTYFGSY
601  YMGREFPHKE RLLKHSRGGGA LQ*

```

ORF128ng and ORF128-1 show 95.7% identity in 622 aa overlap:

```

60  orf128-1.pep  MQAVRYRPEIDGLRAVAVLVSMIFILANNRWLPGGFLGVDIFFVISGFLITGAILSEIQNG
orf128ng       MQAVRYRPEIDGLRAVAVLVSMIFILANNRWLPGGFLGVDIPFVISGFLITNIILSEIQNG
orf128-1.pep  SFSFRDFTTRIKRIYPATIAAVSLASVIASQIFLYEDFNQMRKTVELSAVFLSNLYLGF
65  orf128ng     SFSFRDFTTRIKRIYPATIAAVSLASVIASQIFLYEDFNQMRKTIELSTVFLSNLYLGF

```

	orf128-1.pep	QQGYFDLSADENFVLHIWSLAVEEQYLLYPLLLIFCCKKTKSLRVLNRNIIILFLILTA
5	orf128ng	RLGYFDLSADENFVLHIWSLAVEEQYLLYPLLLIFCCKKTKSLRVLNRNIIILFLILTA
	orf128-1.pep	SSFLPSGYFTDILNQPNYYTLRFPPELLAGSLAVYGTQNGRRQTANGKRLQLLSLC
	orf128ng	SSFLPAGFYTDILNQPNYYTLRFPPELLVGLSLAVYGTQNGRRQTENGKRLQLLSLC
10	orf128-1.pep	FGALLACLFLVIDKHNPFIGMTLLPCLLTALLIRSMQYGTLPTRILSASPIVFGKISY
	orf128ng	FGALLVCLFLVIDKHDPFIGMTLLPCLLTALLIRSMQYGTLPTRILSASPIVFGKISY
15	orf128-1.pep	SLYLYHWIFIAFAHYITGDKQLGLPAVSAVAALTAGFSLLSYLLIEQPLRKRMTEFKAF
	orf128ng	SLYLYHWIFIAFAHYITGDKQLGLPAVSAVAALTAGFSLLSYLLIEQPLRKRMTEFKAF
	orf128-1.pep	FCLYLAPSLILVGYNLYARGILKQEHRLPLPGAPLAAENHFFETVTLTGDSSHAGHLRGFL
20	orf128ng	FCLYLAPSLMLVGYNLYSRGILKQEHRLPLPGTVAENHFFETVTLTGDSSHAGHLRGFL
	orf128-1.pep	DYVSGREGWKAKILSLDSECLVWVDEKLADNPLCRKYRDEVEKAEAVFIAQFYDLRMGGQ
25	orf128ng	DYVSGREGWKAKILSLDSECLVWVDEKLADNPLCRKYRDEVEKAEAVFIAQFYDLRMGGQ
	orf128-1.pep	PVFRFEAQSLIFPGFARFRETVKRIAARKVYVFNANNTSISRSPREEKLRKFAANQYL
	orf128ng	PVFRFEAQSLIFPGFKARFRETVKRIAARKVYVFNANNTSISRSPREEKLRKFAANQYL
30	orf128-1.pep	RPIQAMGDIGKSNQAVFDLIKIDPNVHWVDAQYLPKNTVEIHGRYLYGQDQHLTYFGSY
	orf128ng	RPIRAMGDIGKSNQAVFDLVKIDPNVHWVDAQYLPKNTVEIHGRYLYGQDQHLTYFGSY
35	orf128-1.pep	YMGREFHKKHERLLKSSHGGLQX
	orf128ng	YMGREFHKKHERLLKSSRGGLQX
		610 620

In addition, ORF218ng shows homology to a hypothetical *H. influenzae* protein:

40	sp P43993 Y392 HAEIN HYPOTHETICAL PROTEIN HI0392 >gi 1074385 pir B64007
	hypothetical protein HI0392 - Haemophilus influenzae (strain Rd KW20)
	>gi 1573364 (U32723) H. influenzae predicted coding region HI0392 [Haemophilus
	influenzae] length = 245
	Score = 239 bits (604), Expect = 3e-62
45	Identities = 124/225 (55%), Positives = 152/225 (67%), Gaps = 1/225 (0%)
	Query: 38 VDIFFVISGFLITWIIISLSEIQNSFSFRDFYTRIKRIYFXXXXXXXXXXXXXXFLYE 97
	+DIFFVISGFLIT II++EIQ SFS + FYTRIKRIYF F+Y
	Sbjct: 1 MDIFFVISGFLITGIIITEIQNSFSLKQFYTRIKRIYFPAFITVMALVSFTASAFIYFN 60
50	Query: 98 DFNQMRKTIELSTVFLSNIIYLGFRLLGYFDLSADENFVLHIWSLAVEEQXXXXXXXXXIFC 157
	DFN++RKTIEL+ FLN YLG GFYDLSA+ENFVLHIWSLAVE Q I
	Sbjct: 61 DFNKLRKTIELATAFLSNFYLGTLGQYFDLSANENFVLHIWSLAVEGQYLLIFPLILILA 120
55	Query: 158 YKTKSLRVLNRNIIILFLIITASSFLPAGFYTDILNQPNYYTLRFPPELLVGLSLAV 217
	YKK ++VL I++ILF IL A+SF+ A FY ++LQPN YYLS LRFPPELLVGLSLA+
	Sbjct: 121 YKKFREVKVFIITILILFILLATSFVSANFYKEVLHQPNIIYLSNLRFPELLVGLSLAI 180
	Query: 218 YGTQNGRRQTENGKRLQLSLCFLGALLVCLFLVIDKHDPFIGIT 262
	Y N + Q +L+L L CLF+++ + FIGIT
60	Sbjct: 181 YHNLN-KVQLSKQVNNILAILSTLLLFSCFLFMNNNIAFIGIT 224

This analysis, including the identification of several putative transmembrane domains, suggests that these proteins from *N. meningitidis* and *N. gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 99

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 835>:

```

1  ..ATTATTACG AATACCGCTG GATGTTTCTT TACGGCGCAC TGAACACCTT
5  51  GGGCGTGAAG GTGCTGGCAA C.GCGGGCGG TTGGTATTG GTCTGTGTGT
101 TGGCGTTGCG GCGCCTGATT CACTTGGAAA AAGCCGGTGC GCGGATGGCG
151 GTGCTGGCGT GGGCGTTGCG TAAAGTTTCG CTGCTGTATG TTACGCTGTT
201 CCGGGGTACG CCGCTGTGTTG TGCAGATTGT GATTTGGGCG TATGTGTGTT
251 TCCGTTTTTT CGTC..

```

This corresponds to the amino acid sequence <SEQ ID 836; ORF129>:

```

10 1  ..IIYEYRWMFL YGALTTLGLT VVAXAGGSVL GLLALARLI HLEKAGAPMR
51 51  VLAWLARKVS LLYVTLFRGT PLFVQIVINA YWVFFFFV..

```

Further work revealed the complete nucleotide sequence <SEQ ID 837>:

```

1  ATGGATTTCG GTTTTGACAT TATTACGAA TACCGCTGGA TGTTCCTTA
15 51  CGGCGCACTG ACGACCTTGG GGCTGACGGT CGTGGCAACG GCGGGCGGTT
101 CGGTATTGGG TCTGTTGTGT GCGTTGGCGC GCCTGATTCA CTTGGAAGAAA
151 GCGGTTGCGC CGATGCGCGT GCTGGCGTGG GCGTTGGTGA AAGTTTGGCT
201 GCTGTATGTT ACGCTGTTC GGGGTACGCC GCTGTTTGTG CAGATTGTGA
251 TTTGGCGGTA TGTGTGGTT CCGTTTTCG TCCATCCTTC AGACGGCATT
301 TTGCTCAGCG GCGAGGCGCG AATCGCGCTG CGTCGCGGAT ACGGGCGGCT
20 351 GATTGCGCGT TCTTGGCAC TGCATGCCAA CTCGGGGGCG TATATCTGTG
401 AGATTTTTCG CCGGGGCATC CAGTCTATAG ACAAGAGGAC GATGGAGGCG
451 GCGGTTTCTT TGGGGCTCAG CTATCGCGAG GCGATGCGCT AGTGATTTCT
501 GCGGAGGCA TTGGGCGGCA TGCCTGCGCC TTTGGGAGC GAGTTCATCA
551 CGCTCTTGAA AGACAGCTCG CTGCTGTGCG TCATTGCTGT GCGGAGATTG
25 601 GCGTATGTTT AGAATACGAT TACGSCCGCG TATTGCGTTT ATGAGAACC
651 GCTTTACACC GTGCGCCCTGA TTTATCTGTT GATGAGCAT TCTCTAGGCT
701 GGATATTCTT GCGTTTGGAA AAACGTTACA ATCCGCAACA CCGCTGA

```

This corresponds to the amino acid sequence <SEQ ID 838; ORF129-1>:

```

1  MDRFEDIIFE YRWMFLYGAL TTLGLTVVAT AGGSVLGLLL ALARLIHLEK
30 51  AGAPMRVLAW ALRKVSLLYV TLRGTPLFV QIVIAWVWF PFFVHPSDGI
101 LVSGEAAIAL RRGYGLPIAG SLALIANSGA YICEIFRAGI QSIDKGQMEA
151 ARSLGLTYFQ AMRYVILPQA LRRMLPPLAS EFITLLKDSS LLSVIAVAEL
201 AYVQNTITGR YSYVEEPLYT VALIYLLMTT FLGWIFLRLE KRYNPQHR*

```

Computer analysis of this amino acid sequence gave the following results:

35 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF129 shows 98.9% identity over a 88aa overlap with an ORF (ORF129a) from strain A of *N.meningitidis*:

```

10 20 30 40 50
orf129.pep  IIYEYRWMFLYGALTTLGLTVVAXAGGSVLGLLLALARLIHLEKAGAPMRVLAW
40 10 20 30 40 50 60
orf129a  MDRFEDIIFEYRWMFLYGALTTLGLTVVATAGGSVLGLLLALARLIHLEKAGAPMRVLAW
10 20 30 40 50 60
60 70 80
orf129.pep  ALRKVSLLYVTLFRGTPLFVQIVIAWVWF PFFVHPSDGI
45 10 20 30 40 50 60
orf129a  ALRKVSLLYVTLFRGTPLFVQIVIAWVWF PFFVHPSDGI LVSGEAAIALRRGYGLPIAG
70 80 90 100 110 120
50 orf129a  SLALIANSGAYICEIFRAGIQSIDKGQMEARSLGLTYFQAMRYVILPQALRRMLPPLAS
130 140 150 160 170 180

```

The complete length ORF129a nucleotide sequence <SEQ ID 839> is:

```

1  ATGGATTTCG GTTTTGACAT TATTACGAA TACCGCTGGA TGTTCCTTA
51 51  CGGCGCACTG ACGACCTTGG GGCTGACGGT CGTGGCAACG GCGGGCGGTT

```

-460-

5
10
15
20
25
30
35
40
45
50
55
60
65
70

CGGTATTGGG TCTGTTGTTG GCSTTGGCGC GCCTGATTCA CTTGGA AAAA
GCGCGTTCGC CGATGCGCGT GCTGGCGGTG GCGTTGCGTA AGGTTTGCGT
GCTGTATGTT ACCTGTGTTCC GGGGTACGCC GCTGTTTGTG CAGATTGTGA
TTTGGGCGTA TGTGTGTTT CCGTTTTCG TCCATCCTTC AGACGCGATT
TTGGTTAGCG GCGAGGCGCG AATCGCGGTG CCGCGCGGAT ACGGGCGCGT
GATTTCGCGT TCTTTGGCAC TGATCGCCAA CTCGGGGGCG TATATCTGTG
AGATTPTCCG CCGCGGCATC CAGTCTATAG ACMAAGGACA GATGAGGCGG
GCGGCTCTT TGGGCTGAC CTATCCGCG GCGATGCGCT ATGTGATTCT
GCGCGAGGCA TTGCGCGGTA TGCTGCGCGG TTTGGCGAGC GAGTTCATCA
CGCTCTTGAA AGACAGCTCG CTGCTGTGCG TCATTGCTGT GCGCGAGTTG
GCGTATGTTT AGAATACGAT TAAGGCGCGG TATTGCGTTT ATGAGAGAAC
GCTTTACACC GTCGCCGTGA TTTATCTGTT GATGACGACT TTCTTAGGCT
GGATATTCTT GCGTTTGGA AACCCTTACA ATCCGCAACA CCGCTGA

This encodes a protein having amino acid sequence <SEQ ID 840>:

15
20

1 MDRFDIIYE YRWMFLYGAL TTLGLTVVAT AGGSVLG LLL ALARLIHLEK
51 AGAPMRVLAW ALRKVSLIYV TLRGTPLFV QIVWAYVWF PFFVHPSDGI
101 LVSSEAIAL RRGVPLIAG SLALIANSGA YICEIFRAGI QSTDKQMEA
151 ARSLGLTYPQ ANRYVILPQAL LRMLPPLAS EFTTLKDS LLSVIAVAEL
201 AYVQNTITGR YSVYEPLYT VALIYLLMT FLGWIFLRLE KRYNPQHR*

20 ORF129a and ORF129-1 show 100.0% identity in 248 aa overlap:

25
30
35
40

orf129a.pep MDRFDIIYEYRWMFLYGALTTLGLTVVATAGGSVLGLLLALARLIHLEKAGAPMRVLAW
|||||
orf129-1 MDRFDIIYEYRWMFLYGALTTLGLTVVATAGGSVLGLLLALARLIHLEKAGAPMRVLAW
|||||
orf129a.pep ALRKVSLIYVTLFRGTPLFVQIVWAYVWFPFFVHPSDGI LVSSEAIALRRGVPLIAG
|||||
orf129-1 ALRKVSLIYVTLFRGTPLFVQIVWAYVWFPFFVHPSDGI LVSSEAIALRRGVPLIAG
|||||
orf129a.pep SLALIANSGAYICEIFRAGIQSIDKQMEARSLGLTYPQAMRYVILPQALRMLPPLAS
|||||
orf129-1 SLALIANSGAYICEIFRAGIQSIDKQMEARSLGLTYPQAMRYVILPQALRMLPPLAS
|||||
orf129a.pep EFTTLKDSLSVIAVAELAYVQNTITGRYSVYEPLYTVALIYLLMTFLGWIFLRLE
|||||
orf129-1 EFTTLKDSLSVIAVAELAYVQNTITGRYSVYEPLYTVALIYLLMTFLGWIFLRLE
|||||
orf129a.pep KRYNPQHRX
|||||
orf129-1 KRYNPQHRX
|||||

Homology with a predicted ORF from *N. gonorrhoeae*

ORF129 shows 98.9% identity over a 88 aa overlap with a predicted ORF (ORF129ng) from *N. gonorrhoeae*:

45
50

orf129.pep IIEYRWMFLYGALTTLGLTVVATAGGSVLGLLLALARLIHLEKAGAPMRVLAW 54
|||||
orf129ng MDRFDIIYEYRWMFLYGALTTLGLTVVATAGGSVLGLLLALARLIHLEKAGAPMRVLAW 60
|||||
orf129.pep ALRKVSLIYVTLFRGTPLFVQIVWAYVWFPFFV 88
|||||
orf129ng ALRKVSLIYVTLFRGTPLFVQIVWAYVWFPFFVILHTAFLGNAMQSRVRVPOKRWIAG 120

An ORF129ng nucleotide sequence <SEQ ID 841> was predicted to encode a protein having amino acid sequence <SEQ ID 842>:

55

1 MDRFDIIYE YRWMFLYGAL TTLGLTVVAT AGGSVLG LLL ALARLIHLEK
51 AGAPMRVLAW ALRKVSLIYV TLRGTPLFV QIVWAYVWF PFFVILHTAF
101 LGNAMQSRV VPOKRWIAG SLELNQPRG RKTREGFPPG ESNLGTFRN
151 PLSMGQRFP GCENWYPPQN FTIK*

Further work revealed the following gonococcal sequence <SEQ ID 843>:

1 ATGATTTTC gtTTTGACAT TATTATAcgaA TACCGCTGGA TGTTTCTTTA

-461-

5 251 CGGCGCACTG Acgacettgg ggetgaacgtt cgtggcgacg gCGGCGGGTT
 101 CGGtattggG TCTGTTGTGT GCGTTGGCGC GCGTGTATCA CTTGGAAAAA
 151 GCGGTGCGCG CGATGCGCGT GCTGGCGTGG GCGTTGCGTA AGGTTTCTGT
 201 GCTGTACGTT ACCCTGTTCC GGGGTACGCC GCGTGTTCGT CAGATTGTGA
 251 TTTGGGCGTA TGTGTGTGTT CGGTTTTTTC TCATCCTCTG AGACGGCATT
 301 TTTGTCAGCG GCGAGGCGGC AATCGCGCTG CCGTGGCGAT ACGGGCGCGT
 351 GATTGCGCGT TCTTTGGCAC TGATCGCCAA CTCGGGGGCG TATATCTGTG
 401 AGATTTTCCG CCGCGGCATC CAGCTCATAT ACAAGAGACA GATGGAGGCG
 451 GCGTCTCTT TGGACTGAC CTATCGCGAG GCGATCGCCT ATGTGATTCT
 10 501 GCGCAGCA TCGCGCGCTA TCGTGGCGCG TTTGGCGAGC GAGTTCATCA
 551 CCGTCTTGA AGACAGCTCG CTGCTGCGG TCATTGCTGT GCGCGAGTGT
 601 GCGTATGTT AGAATACGAT TACGGGCGCG TATTCGTTT ATGAAGAACC
 651 GCTTTACACC GCGCGCCTGA TTTATCTGTT GATGACGACT TTCTAGSGCT
 701 GGATATTCCCT GCGTTTGGAA AAACGTTACA ATCCGCAACA CCGCTGA

15 This corresponds to the amino acid sequence <SEQ ID 844; ORF129ng-1>:

1 MDRFRDIIYE YRWMFYGLAL TTLGLTVVAT AGGSVLGLLL ALARLIHLEK
 51 AGAPMRVLAW ALRKVSLLYV TLRGTPLFV QIVIVAYVWF PFFVHPSDGI
 101 LVSGEAIAL RRGYGLPIAG SLALIANSGA YICEIFRAGI QSIDKGQMEA
 151 ARSLGLTYPQ AMRYVILPQA LRRLMPLPAS EFITLLKDSS LLSVIAVAEL
 201 AVYQNTITGR YSYVEEPLYT VALIYLLMTT FLGWIFLRL KRYNPQHR*

ORF129ng-1 and ORF129-1 show 99.2% identity in 248 aa overlap:

orf129-1.pep MDRFRDIIYEYRWMFYGLALTTLGLTVVATAGGSVLGLLLALARLIHLEKAGAPMRVLAW
 25 orf129ng-1 MDRFRDIIYEYRWMFYGLALTTLGLTVVATAGGSVLGLLLALARLIHLEKAGAPMRVLAW
 orf129-1.pep ALRKVSLLYVTLFRGTPLFVQIVIVAYVWFPFFVHPSDGLVSGEAIALRRGYGLPIAG
 orf129ng-1 ALRKVSLLYVTLFRGTPLFVQIVIVAYVWFPFFVHPSDGLVSGEAIALRRGYGLPIAG
 30 orf129-1.pep SLALIANSGAYICEIFRAGIQSIDKGQMEAAARSLGLTYPQAMRYVILPQALRRMLPLPAS
 orf129ng-1 SLALIANSGAYICEIFRAGIQSIDKGQMEAAARSLGLTYPQAMRYVILPQALRRMLPLPAS
 35 orf129-1.pep EFITLLKSSLLSVIAVAELAYVQNTITGRYSYVEEPLYTVALIYLLMTTFLGWIFLRL
 orf129ng-1 EFITLLKSSLLSVIAVAELAYVQNTITGRYSYVEEPLYTVALIYLLMTTFLGWIFLRL
 orf129-1.pep KRYNPQHRX
 40 orf129ng-1 KRYNPQHRX

In addition, ORF129ng-1 is homologous to an ABC transporter from *A. fulgidus*:

2650409(AE001090) glutamine ABC transporter, permease protein (glnP)
 [Archaeoglobus fulgidus]Length = 224
 Score = 132 bits (329), Expect = 2e-30
 45 Identities = 86/178 (48%), Positives = 103/178 (57%), Gaps = 18/178 (10%)
 Query: 65 VSLLYVTLFRGTPLFVQIVIVAYVWFPFFVHPSDGLVSGEAIALRRGYGLPIAGSLAL 124
 +SVY + RQPL VQI+I +F +GI + E A G +AL
 50 Sbjct: 58 ISTAYVEVIRGTPLVQIILI-----VYFLGPAIGNLQPEFA-----GIAL 99
 Query: 125 IANSYAYICEIFRAGIQSIDKGQMEAAARSLGLTYPQAMRYVILPQALRRMLPLASEFTT 184
 SGAYI EI RAGI+SI GQMEA SLG+TY QAMRYVI POA R +IP L +EFI
 Sbjct: 100 SICSGAYIAEIVRAGIESIPGQMEAAARSLGMTYLAQMYRVIFPQAFRNILPALNGEFTA 159
 55 Query: 185 LLIKSSLLSVIAVAELAYVQNTITGRYSYVEEPLYTVALIYLLMTTFLGWIFLRLKRR 242
 LLIKSSLLSVI+ EL V I P AL YL+MT L + +K+
 Sbjct: 160 LLIKSSLLSVISIVELTRVGRQIVNTTNAWTFPLGVALFYLMMTTFLSLRVAYSQKK 217

This analysis, including the identification of transmembrane domains in the two proteins, suggests that the proteins from *N. meningitidis* and *N. gonorrhoeae*, and their epitopes, could be useful
 60 antigens for vaccines or diagnostics, or for raising antibodies.

Example 100

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 845>:

```

1  ..CTGAAAGAAT GCCGTCGAA AGACCGCTGTT TTTATTCCAA ATATCGTTTA
5  51  TAAGAACATC GCCATTACTT TCCTGCTCTT GCACGCCGCC GCCGAACCTT
101  GGCTGCCGCC GCAACGCCGC GGTTTACC GCCTCGCCGT CGGCTTCATC
151  CTGCTCGCCA AGCTGGCGTA gCTTCACCAT CACGAACCTT TACGTAAACA
201  cTAGTCCGCC ACTTATTACy TGCCTCAACT CTTTGCCGCC GCAGgcTAgt
251  TTGTGGACAG GCGCGCGGwA ATTACAAAGC CTGGCCGCTy CCGCGCCCTT
301  GCACCTGATT ACCCTCGGCg GATATAGGG CGCGGTGATg ATGCGTGGG
351  TGACCCCGCG ACTGTGGAC AGCGGCTTTA CCAACTCGA CTACCCCAAA
401  CTCTGCCGCA TTGCGGTGCC CATCGCTTTC GCGCGCGCGC TCTCGCGCGC
451  TTTCTTG+TG AACGTGAACC CG+TATTTT CATTACCGTT CCGTGGATTC
501  TGACCGCGGC CGTATTGTA CTGTATCTTT TC+CGTTTAT ACCGATATT
551  CGGCGCAATG CGTTTACAGA CGATCCGGAr TAr

```

15 This corresponds to the amino acid sequence <SEQ ID 846; ORF130>:

```

1  ..LKECRLLKDFV FIPNIVYKNI AITFLLHAA AELWLPAQTA GFTALAVGFI
51  LLAKLRELHH HELLRKHVYR TYLLQLFAA AGSLWTGAAX LQNLPASAPL
101  HLITLGNMGM GVMVWLTAG LWHSGFTKLD YPKLCRIAPV ILFAAAVRSRA
151  FLXNVNPEFF ITVFAILTAA VFVLYLFXFI PIFRANATFD DPE*

```

20 Further work revealed the complete nucleotide sequence <SEQ ID 847>:

```

1  ATGCGCCCGT TTTTGTCCG GCGCGGGTG CTGCCCACAT TCGGTGCGCT
51  GGTGTTTTTC ATCAACCCCG GTGCCATCGT CCTGCACCGC CAAATTTTCT
101  TGGAACTTAT GTCGCCGCGC GCRPACGGCG GTTTTTTACG TCGGCTTTTG
151  TTGAACTTGA CCGGTTTTTC GGGTAACTGC AAACCTGTGC CGACTTTGAT
201  GCGGCATTA TTGCTGCCG CATCCGCTAT ACTGCCCTTT TCGCGCAAA
251  CTGCTCTGTT TTTGCTGCC GCCTATTGGC TGGTGTTCGT GCTGTTCTGC
301  GCGCGGCTGA TTGGCTAGA CGGAACACC GACAACTTCG CCTGTCTAAT
351  GTTACTTGCC GCGTCTACTG TTTTTCAGAC GGCAATATGC CTGACGCGCG
401  ATTTGAACCT GTTGGCGGCG CAAGTGCATC TAAATATGGC GCGGCTGATG
451  TTCGTATCCG TCGCGGCTCAG TATTTCTTTG GCGCGGGAAG CCTGAAAGA
501  ATGCGGCTCTG AAAGACCCCT TTTTATTTC CAAATATCGT TATAAAAAACA
551  TCGCATTTAC TTTCTGTCTG TTGCACGCGC CGCGCGAATC TTGCTTGCC
601  GCGCAAAACG CCGGTTTTAC CGCGCTCGCC GTCCGCTTCA TCCTGCTCGC
651  CAAGCTGCGT GAGCTTCACC ATCAGAACT CTTACGTAAA CACTACGCTCC
701  GCACATTATA CTTGCTCCAA CTTCTTGCCG CGCAGGCTA TTGTGGACA
751  GCGCGGCGCA AATTACAAAA CCGCGCCGCC TCGCGGCCCC TCACCTGAT
801  TACCCTCGCG GGCATGATGG GCGCGTGAT GATGGTGGG CTGACCGCGC
851  GACTGTGGCA CAGCGGCTTT ACCAACTCG ACTACCCCAA ACTCTGCCCG
901  AITGCGGTCC CCACTCTTTT CGCGCGCGCC GCTCGCGCGC CTTTCTTGAT
951  GAACGTGAAC CCGATATTTT TCATTACCGT TCCTGGGATT CTGACCGCGC
1001 CCGTATTGCT ACTGATCTT TACAGCTTTA TACCGATATT TCGGCGCAAT
1051 CGGTTACAG ACGATCCGGA ATAA

```

This corresponds to the amino acid sequence <SEQ ID 848; ORF130-1>:

```

1  MREFFVGAAY LAILGALVFF INPGAIVLHR QIFLEMLPEA AYGGLTAAAL
45  LDWTGFGSNL KPVATLMAAL LLAASAILFF SPQTASFFVA AYNLVLLFLC
101  ARLIWLDRNT DNFAILLMLL AFTVFQTAYA VSGDLNLLRA QVHLNMAAVM
151  FVSVRVSIIL GAELAKRECL KDPVFIPNIV YKNIATITFL LHAARLEWLF
201  AQTAGFTALA VGFILLAKLR ELHHHELARK HYVRTYLLQ LEAAGYIWT
251  GARLQNLEA SAPHLITLGL GMMGGVMVMV LTAGLWHSFG TKLDYFKLCR
301  IAVELLFAAA VSRALMNVN FIEFITVPAL LTAAVVLYLV FTPIPIFRAN
351  AFTDDPE*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF130 shows 94.3% identity over a 193aa overlap with an ORF (ORF130a) from strain A of *N.*

55 *meningitidis*:

[illegible]

25	1	ATGCGGCGCGT	TTTTCTGTGG	CGGCGGGGTG	CTTGCACATAC	TCGGTGGCGCT
	51	GGTGTTTTC	ATCAACCGCG	TGCGGACATG	CTGCACGCGC	CAAAATTTTCT
	101	TGCACTATTC	CTGCAGCGCG	GACATACGGCG	GTTTCTTGAC	TGCGGCTTTTG
	151	TGGACTTGA	CGGGTTTTC	GGGTATGCGT	AAACCGTCTG	CGACTTTGAT
30	201	GGCGCGATT	TGTGCTCGCG	CATCGCGAT	ATCGCCCTTT	TGCGCGCAAA
	251	CTGCGCTGT	TTTGTGTCGG	GCGATATGCG	TGGCTGTGTG	CGTGTCTCGC
	301	CGCGCGCTGA	TGTTGCTAGA	CGAAGAACG	GACACATCTG	CCCTGCTAAT
	351	GTTACTTGGC	GGTGTCACTG	TTTTCACAG	GACATATGCG	TGCGAGCGGG
	401	ATTTGAACTT	TGCGCGCGCG	CAGTGTGCTG	TAAATATGCG	CGCGGTGATG
	451	TTCTGATATG	TGCGGCGCGG	TATGCTGCGG	CTGCAAAATG	CGCGGCGGCG
35	501	ATGCGCTCTG	AAAGACCGAG	TATCTATGCG	CAATGTCTCG	TATAAAAACA
	551	TGCGCATATC	CTCTGCTGTC	TGCGCGCGCG	CGCGCGCACT	TGCGCTGCTC
	601	GCGCAACCGC	CGGGTTTATC	CTGCGCTGCG	CTCGGCTTAT	TCCTGCTTGC
	651	CAGCTCGGCT	GAGCTTCAAC	ATACAGCAAT	CTCGCGGAAA	CACATCGTCC
	701	GCACTTATTA	TGCTGCTCAA	CTCTTTCGCG	CGCGAGGCTG	TTTGTGCGCA
40	751	GGCGCGCGCA	ATATACAAAA	CTCGCGCCG	TGCGGCGTGC	TGCACTTGAT
	801	TACCTCTGCT	GGCATGATGG	GACAGCTGAT	TTGCTGTGGG	TGCACTGCGG
	851	GACTCTTGAC	CGCGCGCTTT	ACCAAGCTCT	ATCGACCGAA	ACTCTGCGCG
	901	ATGCGCGCGC	CGCGCGCGCG	TGCGGCGCGG	CTGCGGCGCG	CTCTGCGGCG
	951	GAGCTGAAC	CGGATATTCT	TACATACGCT	CCGCGCATCT	CTCTGCGGCG
45	1001	CGCTGTTCAG	CTTTTACCTG	CTACCATCTG	TACCGATCTT	TGCGGCGGAC
	1051	GCGTTTACAG	ACGATCCGGA	ATAA		

	1	MRFFVFGSNV	LAILGALVFF	LNQSAIVLHR	QTFLEMLPA	AYGGFLTRAL
50	5	LDWFGQAGAL	KPVATINLFF	IPGAAALFSP	SPQAFASVFA	AYKLVLLLPF
	10	ALIRLWLRNT	DFNMLNLIA	AFDVFQVFA	VSQGLMLLRA	QVHNLMAVMV
	15	FVSVRSVILL	GAELAKCEKL	KP2PTEINVV	YKNATITPLL	LHAAAEWLNP
	20	AQATGATLFA	VGFILLAKLR	ELHHEMLLRK	HVYRTYLLQL	LFAAGAYLTK
	25	GAKLQNLPA	SAPHLITLGL	GMGVSGVSGF	PLAGLHSGSF	TKLDYPKLKR
55	30	IAPVPLFAAA	VSRVAMNVNV	P1FFTFVPAI	LTAAVFVLYL	LTFVPIFRAN
	35	ATFDDEP*				

60 orf130a.pep MRPFVFGAAVLA L LGA LVFFNPGA V LHRQFLEIM L PAAYGGFLTAALLDWTGFGSNL
orf130-1 MRPFVFGAAVLA L LGA LVFFNPGA V LHRQFLEIM L PAAYGGFLTAALLDWTGFGSNL

orf130a.pep KPVATLMAALLAASAILFPSPQTASFVVAAYVLVLLFLFCARLWLDRNTDNFALLMLLA
orf130-1 KPVATLMAALLAASAILFPSPQTASFVVAAYVLVLLFLFCARLWLDRNTDNFALLMLLA

65 orf130a.pep KFTVKTQAYAVSGDNL LRACVHLNNAAMFVSVRVSL LGAELKECKRLKDXPVIINVVV

	orfl30-1		AFTVFQTAAYAVSGDLNLLRAQVHLNMAAVMFVSVRVSIILLGAEALKECRKDPVFIPNIV	
5	orfl30a.pep		YKNIAITFLLHAAELWLPAQTAGTSLAVGFILLAKLRELHHHELLRKHYVRTYLLQ	
	orfl30-1		YKNIAITFLLHAAELWLPAQTAGTSLAVGFILLAKLRELHHHELLRKHYVRTYLLQ	
	orfl30a.pep		LFAAAGYLWTGAARKLQNLPAASAPHLHLITLGGMMGGVMMVWLTAGLWHSGETKLDYPKLCR	
10	orfl30-1		LFAAAGYLWTGAARKLQNLPAASAPHLHLITLGGMMGGVMMVWLTAGLWHSGETKLDYPKLCR	
	orfl30a.pep		IADVILFAAAVSRVLMNVNPIFFITVPAILTAAPVLYLLTFVPIFRANAFTDDPE	
	orfl30-1		IADVILFAAAVSRVLMNVNPIFFITVPAILTAAPVLYLLTFVPIFRANAFTDDPE	

15 Homology with a predicted ORF from *N.gonorrhoeae*

ORF130 shows 91.7% identity over a 193 aa overlap with a predicted ORF (ORF130ng) from *N.gonorrhoeae*:

20	orfl30.pep		LKECRKDPVFIPNIVYKNIAITFLLHAA	30
	orfl30ng	LNLLRAQVHLNMAAVMFVSVRVSVLLGTETLKECRKDPVFIPNIVYKNIAITFLLHAA		201
	orfl30.pep		AELWLPAQTAGTALAVGFILLAKLRELHHHELLRKHYVRTYLLQLFAAAGSLWTGAAX	90
25	orfl30ng		AELWLPAQTAGTALAVGFILLAKLRELHHHELLRKHYVRTYLLQLFAAAGYLWTGAAG	261
	orfl30.pep		LQNLPAASAPHLHLITLGGMMGGVMMVWLTAGLWHSGETKLDYPKLCRIADVILFAAAVSRA	150
	orfl30ng		LQNLPAASAPHLHLITLGGMTGGVMMVWLTAGLWHSGETKLDYPKLCRIADVILFAASVRA	321
30	orfl30.pep		FLXNNVXKFFITVPAILTAAPVLYLLTFVPIFRANAFTDDPE	193
	orfl30ng		VLMNVNPIFFITVPEILTAAPVLYLLTFVPIFRANAFTDDPE	364

An ORF130ng nucleotide sequence <SEQ ID 851> was predicted to encode a protein having amino

35 acid sequence <SEQ ID 852>:

	1	MNKEFTHPMR	FFVGAAVLA	ILGALVFFHQ	FRRYHAPPN	FLGTAAAGCI
	51	RRFFDYRFVG	PDGFFRQPET	CRYFDGGVVA	CGCGFIAVET	ATCRIFRRRL
	101	LAVGAVALRL	ADLARRQHRT	LRSVDVTRAF	TVFQTAIVAVS	GDNLNLLRAQV
	151	HLNMAAVMFV	SVRVSVLLGT	ETLKECRKLD	VPFIPNIVYK	NIAITLLHHA
40	201	AELWLPAQT	AGTALAVGF	ILLAKLRELH	HHELLRKHYV	RTYLLQLQFA
	251	AAGYLWTGAA	LQNLPAASAP	LHLITLGMKT	GOVMMVWLT	GLWHSGETKL
	301	DYPKLCRIAV	SILFAASVSR	AVLMNVNPIF	FTVPEILTA	AVFMYLLTFE
	351	VPIFRANAFT	DDPE*			

Further work revealed the following gonococcal DNA sequence <SEQ ID 853>:

45	1	ATGGCGCCGGT	TTTTCGTGG	TGCGGCAGTA	CTTGCCATAC	TGCGTGGGTT
	51	GSTGTTTTTT	ATCAACCCCG	GCGCTATCAT	CTCGACCCGC	CAAAATTTTCT
	101	TGGAACCTAT	GCTCGCGGCT	GCAATACGGG	GTTTTCGATC	TACCGCTTTG
	151	TTGGAACCGA	CGGGTTTTC	AGGCAACCTG	AAACCTGCGC	CTACTTTGAT
	201	GGCGGTGTG	TGCTGTGTG	CGGCTGTTTT	ATTGCGCGTT	TTACCGCAAC
50	251	TTGCGGCATT	TTTTCGTGCG	GCTATTGGC	TGGTGTGTGT	GCTGTCTGCT
	301	GCTCGGCTGA	TTTGCTCGA	CGGCAACACC	GACAACTTGC	CTCTGTTGAT
	351	GTTACTTGCC	GCATTTACCG	TTTTTCAGAC	GGCCTATGCC	GTCAGCGGCG
	401	ATTTGAACCT	ACTGCGCGCG	CAAGTCGATT	TGAATATGCC	GCGGCTCATG
	451	TTGCTATCCG	TGCGGCTCAG	CGTCTCTTTG	GGCACGGAAA	CCCTGAAAGA
55	501	ATGCGGCTCG	AAAGACCCCG	TATTCATCCC	CAACGTTATC	TATAAAAAACA
	551	TGCGCATCAC	CTGCTGCTGT	CACGCGCGCG	CGGAACCTTGT	GTCGCGCGCG
	601	CAAGCGCGCG	GTTTACTACTG	TGCTGCGCGC	GCGCTCATCC	TGCTCGCGCA
	651	CTCGCGGAA	CTGCAACCTC	ACCAACTCTT	ACGCAACCTC	TACGTCGCGA
	701	CTTATTACCT	GCTCCAGCTC	TTTGCGCGCG	CAGGTTATCT	TGCGACAGCG
60	751	GCGCGGAAAC	TGCAAAACCT	GCGCGCGCTC	GCGCGCGCTC	ACCTGATTAC
	801	CCTCGCGCGC	ATGACGGGCT	GCGTATGAT	GCGTGGCTGT	ACTGCGCGAC
	851	TGTGGCACAG	CGGCTTTTAC	AACTCGACT	ACCGGAACT	CTGCGCGCAT

```

901 GCGCTGCCA TCCTTTTCGC CTCGCGCGT TCGCGCGCTG TTTTAATGAA
951 CGTGAATCCG ATATTCTTCA TCACCGTCC CGAGATTCTG ACCGCGCGCG
1001 TGTTCATGCT TACCTGCTG ACGTTCGTAC CGATTTTTCG AGCGAACGCG
1051 TTTACAGACG ATCCGGAATA A

```

5 This corresponds to the amino acid sequence <SEQ ID 854; ORF130ng-1>:

```

1 MRPFVGAARV LAILGALVFF INPGAIIILHR QIFLEMLPFA AYGGFLTTAL
51 LDRTGFSGNL KPAATLMARV LIVAVALLPF LPQLAAFFVA AYWLVLLEFC
101 AWLIWLDNRT DNFALLMLLA APTVQTAYVA VSGDLNLLRA QVHLNMAVNV
151 FVSVRVSVLL GFTLKLRLR KDPVTFPNTV YKNIAITLLL HAARELWLPFA
201 QTAGFTALAV GFTLKLRLR LHHHELLRKH YVRYTYLLQL PAAAGYLWTG
251 AAKLQNLPAF APLHLITLGG MTCGVMVMWL TAGLWHSGET KLDYPKLCRI
301 AVSILFASAV SRAVLNVMNF IFFITVPEIL TAAVFNLYLL TFVPIFRANA
351 FTDDEP*

```

ORF130ng-1 and ORF130-1 show 92.4% identity in 357 aa overlap:

```

15 orf130-1.pep MRPFVGAARV LAILGALVFF INPGAIIILHR QIFLEMLPFA AYGGFLTTAL LDRTGFSGNL
    orf130ng-1 MRPFVGAARV LAILGALVFF INPGAIIILHR QIFLEMLPFA AYGGFLTTAL LDRTGFSGNL
20 orf130-1.pep KPVATIMAAALLAASAILPFPSTQASFFVAAYVLWLLFCARLIWLDRNTDNFALLMLLA
    orf130ng-1 KPAATLMARV LIVAVALLPF LPQLAAFFVAAYVLWLLFCARLIWLDRNTDNFALLMLLA
25 orf130-1.pep AFTVFQYAVSVSGDLNLLRAQVHLNMAVMFVSVRVSVILGAEALKECRCLKDPVTFIPNV
    orf130ng-1 AFTVFQYAVSVSGDLNLLRAQVHLNMAVMFVSVRVSVILGAEALKECRCLKDPVTFIPNV
30 orf130-1.pep YKNIAITFLLHAAELWLPQAQTAGFTALAVGFILLAKLRELLHHELLRKHRYVRYTYLLQ
    orf130ng-1 YKNIAITFLLHAAELWLPQAQTAGFTALAVGFILLAKLRELLHHELLRKHRYVRYTYLLQ
35 orf130-1.pep LFAAAGYLWTGAARKLQNLPAFAPLHLITLGGMTGVMVMWLTAGLWHSGETKLDYPKLCR
    orf130ng-1 LFAAAGYLWTGAARKLQNLPAFAPLHLITLGGMTGVMVMWLTAGLWHSGETKLDYPKLCR
40 orf130-1.pep IAVPILFAAAVSVRAFLNVMNPIFFITVPAILTAAVFVLYLTFPIPRANAFTDDEP
    orf130ng-1 IAVSILFASAVSRAVLNVMNPIFFITVPEILTAAVFNLYLLTFVPIFRANAFTDDEP

```

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

40 Example 101

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 855>:

```

1 ATGGAAATTC GGGCAATAAA ATATACGGCA ATGGCTGCGT TGCTTGCAAT
51 TAGCGTTGCA GCGTGCOCGGT TGGCGGGGTG GTATGAGTGT TCGTCCCTCA
101 CGCGTGGTGT TAGCCGAGAA AAACGCGGCTG CAGTCGATTG TTGGAGTATT
151 GCGCGCGAGA GTCCGCGGCTC TTGAGGGGAC TAGCAGATAC CGCTTCAGAA
201 CGGCAATAGT TCGCTCAGGG CAAACGGAATA TGAATCCGCA CAACATCTT
251 ACTTTTACAG GAAATATAGG AAGTTTGAAG C.TGCGGCGT GGATTTGGGT
301 ACGCGTGACG GCAACCTTT GATTGAGAAG TTCAACAGG GAGGATTGTA
351 CTGCTTGGAA AAG..

```

50 This corresponds to the amino acid sequence <SEQ ID 856; ORF131>:

```

1 MEIRAIKYTA MAALLAFTVA GRLAGWYEC SSLTGWCKPR KPAADIFWDI
51 GGSPSPSLGD YEIPLSDGNS SVRANEYESA QQSIFYRKIG KFEXCGLDWR
101 TRDGKPLIET FKQGGFDCLE K..

```

Further work revealed the complete nucleotide sequence <SEQ ID 857>:

```

55 1 ATGGAAATTC GGGCAATAAA ATATACGGCA ATGGCTGCGT TGCTTGCAAT
    51 TAGCGTTGCA GCGTGCOCGGT TGGCGGGGTG GTATGAGTGT TCGTCCCTCA

```

5

101	CCGGCTGGTG	TAAAGCGAGA	AAACCGGCTG	CCATCGATT	TTGGGATATT
151	GCGGCGGAGA	GTCGCGCGTC	TTTAGGGGAC	TACGAGATC	CGCTTCAGA
201	CGCGAATCGT	TCGTCGATGG	CAACAGGATA	TGAATCCGGA	CAACATCTT
251	ACTTTACAG	GAATAATGTT	GATTTGAAG	CTCCGCGGCT	GGATTGGCT
301	ACGGTCGACG	CAAAACTCTT	GATTTGACG	TTCAACAGAG	GAGGATGAT
351	CTGCTTGCAA	AAGCAGGGGT	TGCGGCGCAA	CGGTCTGTCC	GAGCGCGTCC
401	GATCGTAA				

This corresponds to the amino acid sequence <SEQ ID 858; ORF131-1>:

10 1 MEIRAIKYTA MAALLAFTVA GCRLAGWYEC SSLTGWCKPR KPAADIFWDI
51 GGESPPSLGD YEIPLSDGNR SVRANEYESA QQSYFYRKIG KFEACGLDWR
101 TRDGKPLIET FKOGGFDCLE KCGLRNGLS ERVRW*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF131 shows 95.0% identity over a 121aa overlap with an ORF (ORF131a) from strain A of *N.*

15 *meningitidis*:

		10	20	30	40	50	60
	orf131.pep	MEIRAITYTAMAALLAFTVAGCRLAGWECSSLTGWCKFRKPAADFWDWIGGESPPSLG					
20	orf131a	MEIRAITYTAMAALLAFTVAGCRLAGWECSSLGSGWCKFRKPAADFWDWIGGESPPSLG					
		10	20	30	40	50	60
	orf131.pep	YEIPLSDGNSSVRANEYESAQCSYFYRKIGKGFEXCGLDWRTRDGKPLIETFKQGGFDCLE					
25	orf131a	YEIPLSDGNSSVRANEYESAQCSYFYRKIGKGFEXCGLDWRTRDGKPLIETFKQGGFDCLE					
		70	80	90	100	110	120
		70	80	90	100	110	120
	orf131.pep	K					
30	orf131a	KQGLRRLNGLSERVVWX					
		130					

The complete length ORF131a nucleotide sequence <SEO ID 859> is:

35	1	TACGGAAATTC	GGGCATATAA	ATATACGCCA	GTGGCGCGGT	TGCTTGCATT
	5	TACGGTGTCA	GCGTCGCGGT	TGGCAGGTGTG	ATGTAGTGTT	TGTCCTCGAT
	10	CCGCGTGGTG	TACCGCGAGA	AAATCGTCGC	CGCTCATGTT	TGGGATATAT
	15	CGCGCGGAGA	GTCCCTCCGTC	TTTGTAGAGAC	CATCAGATAC	CGCTTTCAGA
	20	CGGCATTCGT	TCCTCGTAGG	CACAAGUATA	TGATATCCGA	CACAACTATT
40	25	ACTTTACAG	GAATAATGGG	AAGTTTGAAG	TGCTGCTGCT	GGATTTCGCT
	30	ACGCTGCGC	GCAACCTTT	GATTGACAGA	TTTCAACAGC	AAGSTTTTGA
	35	TTTGTGAGA	GCACAGGGT	TGCGCGCGA	CGGTTGTTC	GACGCGCTCT
	40	TGTTGTAA				
	45					

This encodes a protein having amino acid sequence <SEO ID 860>:

45 1 MEIRAIKYTA MAALLAFTVA GCRLAGWYEC SLSGWCKPR KPAAIDFWDI
51 GGESPPSLED YEIPLSDGNR SVRANEYESA QSYFYRKIG KFEACGLDWR
101 TRDGKPLIET FKCEGFCLK KQGLRRNGLS ERVR*

ORF131a and ORF131-1 show 97.0% identity in 135 aa overlap:

```

50      | orf131a.pep      MEIRATKYTAAALAFYVAGCRLAGWYECSSLGSGWCKPKKPAIDFDWIGGSESPSLSD
      |                  |||||
      | orf131-1        MEIRATKYTAAALAFYVAGCRLAGWYECSSLGSGWCKPKKPAIDFDWIGGSESPSLSD
      |                  |||||
      | orf131a.pep      YEIPLSDGNRSVRANYESAQCSYFYRKIGKFEACGLDWRTRDGKPLIETFKGGFDCL
      |                  |||||
      | orf131-1        YEIPLSDGNRSVRANYESAQCSYFYRKIGKFEACGLDWRTRDGKPLIETFKGGFDCL
      |                  |||||
55      | orf131a.pep      KQGLRRNGLSERVRWY
      |                  |||||

```


orf131-1 KQGLRRNGLSERVRW

Homology with a predicted ORF from *N.gonorrhoeae*

ORF131 shows 89.3% identity over 121 aa overlap with a predicted ORF (ORF131ng) from

5 *N.gonorrhoeae*:

```

orf131.pep  MEIRAIKYTAMAALLAFTVAGCRLAGWYECSSLTGWCKPRKPAIDFWDIGGESPPSLGD  60
orf131ng    MEIRVIKYTATAALFAFTVAGCRLAGWYECSSLTGWCKPRKPAIDFWDIGGESPPSLGD  60

10 orf131.pep  YEIPLSDGNRSVRANEYESAQSSYFYRKIGKFEACGLDWRTRDGKPLIETFKQGGFDCL  120
    orf131ng    YEIPLSDGNRSVRANEYESAQSSYFYRKIGKFEACGLDWRTRDGKPLIVERFKQGGFDCL  120

15 orf131.pep  K 121
    orf131ng    KQGLRRNGLSERVRW 134

```

A complete length ORF131ng nucleotide sequence <SEQ ID 861> was predicted to encode a protein having amino acid sequence <SEQ ID 862>:

```

20 1 MEIRVIKYTA TAALFAFTVA GCRLAGWYEC LSLSGWCKPR KPAIDFWDI
51 GGESPLSLED YEIPLSDGNR SVRANEYESA QKSYFYRKIG KFEACGLDWR
101 TRDGKPLIVER FKQGGFDCL KQGLRRNGLS ERVRW*

```

Further work revealed the following gonococcal DNA sequence <SEQ ID 863>:

```

25 1 ATGGAATTC GGGTAATAA ATATACGGCA ACGGCTGCGT TGTTCGATT
51 TACGGTTGCA GCGTCGCGG TGGCGGGGTG GTATGAGTGT TCGTCTTGT
101 CCGGCTGGTG TAAGCCGAGA AAACCTGCCG CCATGAGTTT TTGGGATATT
151 GCGGCGGAGA GtcgctgGTC TTTAGAGGAC TACAGATATC CGCTTTCAGA
201 CGGCAATCGT TCCGTCAGGG CAACAGATA TGAATCCGGC CAAAAATCCT
251 ACTTTTATAG GAAATAGGG AAGTTTGAAG COTGCGGGTT GGATTGGCGT
301 ACGCGTAGCG GCAACCTTT GGTTCAGAGG TTCAAACAGG AAGGTTTCGA
351 CTGTTTGGAA AAGCAGGGGT TGGCGGCGAA CGGCTCTGCC GAGCGCGTCC
401 GATGGTAA

```

This corresponds to the amino acid sequence <SEQ ID 864; ORF131ng-1>:

```

35 1 MEIRVIKYTA TAALFAFTVA GCRLAGWYEC SSLSGWCKPR KPAIDFWDI
51 GGESPLSLED YEIPLSDGNR SVRANEYESA QKSYFYRKIG KFEACGLDWR
101 TRDGKPLIVER FKQGGFDCL KQGLRRNGLS ERVRW*

```

ORF131ng-1 and ORF131-1 show 92.6% identity in 135 aa overlap:

```

40 orf131ng-1.pep MEIRVIKYTATAALFAFTVAGCRLAGWYECSSLTGWCKPRKPAIDFWDIGGESPLSLED
    orf131-1      MEIRAIKYTAMAALLAFTVAGCRLAGWYECSSLTGWCKPRKPAIDFWDIGGESPPSLGD

orf131ng-1.pep YEIPLSDGNRSVRANEYESAQSSYFYRKIGKFEACGLDWRTRDGKPLIVERFKQGGFDCL
orf131-1        YEIPLSDGNRSVRANEYESAQSSYFYRKIGKFEACGLDWRTRDGKPLIETFKQGGFDCL

45 orf131ng-1.pep KQGLRRNGLSERVRW
    orf131-1      KQGLRRNGLSERVRW

```

Based on the presence of a predicted prokaryotic membrane lipoprotein lipid attachment site, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 102

The following partial DNA sequence was identified in *N. meningitidis* <SEQ ID 865>

```

1  ATGAAACACA  TCCATATTAT  CGGTATCGGC  GGCACGGTTA  TGGGCGGGCT
5  51  TCGCGCCATT  GCCAAGAAGG  CGGGGTTTGA  AGTCAGCGST  TCGCAGCGGA
101 AGATGTATCC  CGCGATGAGC  ACCCAGCTCG  AAGCCTTGGG  TATAGACGTG
151 TATGAAGGCT  TCGATGCGCG  TCAGTTGGAC  GAATTTAAAG  CGACGTTTAA
201 CGTTATCGGC  AATGTGCGCA  AGCGCGGGAT  GGATGTGGTT  GAAGCGATT
251 TGAACCTCGG  CTCGCTTAT  ATCTCGGCC  CGCATGGCT  GTCCGAAAC
301 CTGCTGCACC  ATCTTGGCT  ACTGCGAGC  CGCGGAGCG  ACAGCAACAC
10  351 GACCACCGCC  TCCATGCTCG  CATGGCTCTT  GGAATATGCG  GGCTCGCGCG
401 CGGGCTTCTT  TATGCGCGCG  GTACC. GGAA  AATttCGGCG  TTTCGCGCG
451 CTGCGCGCAA  ACGCGCGGCC  AAGACCGGAA  CAGCCCAATC  CGGTTTTTCG
501 TCATCGAAGC  ACGAAGATAC  GACACCGCCT  TTTTCGACAA  ACGTTCTAAA
15  551 TtCGTGCATT  ACGCTCGCG  TACCGCGCTG  TTGAACAATC  TGAATTCGAA
601 CACGCGCGAC  ATCTTTGCGC  ACTTGGCGCG  GATACGACAC  CAGTCCACT
651 ACCTCGTGC  TACCGTGGCG  TCTGAAGGCT  TAATCGTCTG  CAACGCGAGG
701 CAGCAAAAGC  TCGAAATAC  TTTGGACAAA  GGCTGCTGGA  CGCCGTTGGA
751 AAAATTGCGC  ACGGAACACG  GCTGGCA..

```

This corresponds to the amino acid sequence <SEQ ID 866; ORF132>:

```

20  1  MKHHIIGIG  GTFMGGLAAI  AKEAGFEVSG  CDKMYPPMPS  TQLEALGIDV
51  YEGFDAQILD  EFKADVVYIG  NVAKRGMDVV  EAILNLGLPY  ISGFQWLSN
101 VLHHHNVLVG  AGTHGKTTTA  SMLAWVLEYA  GLAPFLIGG  VNGKFRFRFP
151 PAANAAPRPE  QPIAVPRHRS  RRIHRHFRQ  TFXIRALPSA  YRVEQSGIR
201 PRHRLCLRGR  DDEVFLFRA  YRAVXRNLRL  QRTAAKPARY  FQRLLDAGG
25  251 KIRHGTRIA..

```

Further work revealed the complete nucleotide sequence <SEQ ID 867>:

```

1  ATGAAACACA  TCCATATTAT  CGGTATCGGC  GGCACGGTTA  TGGGCGGGCT
5  51  TCGCGCCATT  GCCAAGAAGG  CGGGGTTTGA  AGTCAGCGST  TCGCAGCGGA
101 AGATGTATCC  CGCGATGAGC  ACCCAGCTCG  AAGCCTTGGG  TATAGACGTG
151 TATGAAGGCT  TCGATGCGCG  TCAGTTGGAC  GAATTTAAAG  CGACGTTTAA
201 CGTTATCGGC  AATGTGCGCA  AGCGCGGGAT  GGATGTGGTT  GAAGCGATT
251 TGAACCTCGG  CTCGCTTAT  ATTTCCGGCC  CGCAATGGCT  GTCCGAAAC
301 GTGCTGCACC  ATCATTTGGT  ACTCGGTGTG  CGCGGGACGC  ACAGCAACAC
351 GACCAACGCG  TCCATGCTCG  CATGGCTCTT  GGAATATGCG  GGCTCGCGCG
35  401 CGGGCTTCTT  TATTTGCGCG  GTACCGGAA  ATTTCCGGCG  TTTCGCGCG
451 CTGCGCGCAA  CGCGCGGCCA  AGACCGCAAC  AGCCAATCGC  CGTTTTTCGT
501 CATCGAAGCC  GACGAATACG  ACCCGCCTT  TTTTCGACAA  CGTTCTAAAT
551 TCGTGCAATTA  CGGTGCGCGT  ACCGCGCTGT  TGACCAATCT  GGAATTCGAC
601 CACGCGGACA  TCTTTGCGCA  CTGGGCGCG  ATACAGACCC  AGTTTCACTA
40  651 CCTCGTGCAT  ACGTGGCGGT  CTGAAGGCTT  AATCGTCTCG  AACGGACGGC
701 AGCAAAAGCT  CCAAGATACT  TTGGACAAGG  GCTGCTGAGC  GCGGTGGAA
751 AAATTTCGGCA  CGGAACACGG  CTGGCAGCGC  GGCGAAGCCA  ATGCGGACGG
801 CTGCTTCGAC  GTGTTGCTCG  ACGGCAAAAC  CGCCGAGACG  GTCAAAATGG
45  851 ATTTGATGGG  GAGGCACAC  CGCATGAACG  CGCTCGCGCT  CATTGCGCGC
901 GCGCGTCAIT  TCGGTGTGCA  TATTAGACCC  GCCTGCGAAG  CTTTGGCGCG
951 GTTTAAAAAC  GTCAACAGCC  GGATGGAAT  CAAGGCGACG  GCAAAACGCA
1001 TCACCGTTTA  CGCGACTTC  GCGCACACC  CAGCCGCAAT  CGAAACCGCA
1051 ATTCAAGGTT  TCGCGCAAGC  CTTGCGCGCG  CGCGCATCTC  TCGCGTCTCT
1101 CGAACCGGCT  TCCAAACAGA  TGAAGCTGGG  CACGATGAAG  TCGCGCTCTG
50  1151 CTGTAAGCCT  CAAAGAAGCC  GACCAAGTGT  TCTGTACGC  CGCGCGCGTG
1201 GACTGGGACG  TCGCGGAGC  CCTCGCGCT  TTGGGCGGCA  GGCTGACGCT
1251 CGGCAAAAGC  TTTGATGCTT  TCGTTGCGGA  AATCTGAAA  AACGCGAAG
1301 TAGGCGACCA  TATTTTGGTG  ATGAGCAACG  GCGGTTTCGG  CGGAATACAC
1351 GGAAGCTGCG  TGAAGCTTT  GAGATAG

```

This corresponds to the amino acid sequence <SEQ ID 868; ORF132-1>:

```

1  MKHHIIGIG  GTFMGGLAAI  AKEAGFEVSG  CDKMYPPMPS  TQLEALGIDV
51  YEGFDAQILD  EFKADVVYIG  NVAKRGMDVV  EAILNLGLPY  ISGFQWLSN
101 VLHHHNVLVG  AGTHGKTTTA  SMLAWVLEYA  GLAPFLIGG  VNGKFRFRFP
151 LPQTFRQDFN  SQSFVFVIEA  DEVDYAFDFK  RSKFVHFRFP  TAVLNKLFED
60  201 HADIFALGIA  IQTFPIVLVR  TVPSEGLIVC  NRGQSLQFT  LDGQCTVPE
251 KFTGHEWQGA  GEANADGSFD  VLLDGKTAGR  KWLDLGRHN  RMNALAVIAT
301 ARHVGVDIQT  ACEALGAFKN  VKRRMEIKGT  ANGITYVDYDF  AHHPTAIETT

```

```

351 IQGLRQRVGG ARILAVLEFR SNTMKLGTMK SALPVSLEKEA DQVFCYAGGV
401 DWDVAEALAP LGRLNVGKD FDAFVAEIVK NAEVGDHILV MSNGGFGGIH
451 GKLEALR*

```

Computer analysis of this amino acid sequence gave the following results:

5 Homology with the hypothetical o457 protein of *E. coli* (accession number U14003)

ORF132 and o457 show 58% aa identity in 140 aa overlap:

```

10 Orf132: 4 IHIIGIGGTFMGGGLAAIAKEAGFEVSGDAKMYPFMSQTLEALGIDVYEGFDAAQLDEFK 63
    o457: 3 IHI+GI GTFMGGGLA +AA+ G EV+G DA +YPMST LE GI++ +G+DA+QL+ +
    IHIIGIGGTFMGGGLAARLQGLHEVTGSDANVYPMSTLLEKQGIELIQGYDASQLEP-Q 61

Orf132: 64 ADVYVIGNVAKRGMDVVEAILNLGLFYISGPQWLSENVLHHHWLVGVAGTHGKTTTASML 123
    D+ +IGN RG VEA+L +PY+SGPQWL + VL WVL VAGTHGKTTT M
o457: 62 PDVLIIGNNATRGNFCEAVLEKNIPYMSGPGWLHDFVLRDRWLVAVAGTHGKTTTAGMA 121

15 Orf132: 124 AWVLEYAGLAPGLIGGVXG 143
    W+LE G PGF+IGGV G
o457: 122 TWILEQCQGVKPGFVIGGVPG 141

```

Homology with a predicted ORF from *N. meningitidis* (strain A)

20 ORF132 shows 74.6% identity over a 189aa overlap with an ORF (ORF132a) from strain A of *N.*

meningitidis:

```

25 orf132.pep MKHHIHIIGIGGTFMGGGLAAIAKEAGFEVSGDAKMYPFMSQTLEALGIDVYEGFDAAQLD
    orf132a MKHHIHIIGIGGTFMGGGLAAIAKEAGFEVSGDAKMYPFMSQTLEALGIDVYEGFDAAQLD
    10 20 30 40 50 60
    10 20 30 40 50 60

orf132.pep EFKADVYVIGNVAKRGMDVVEAILNLGLFYISGPQWLSENVLHHHWLVGVAGTHGKTTT
    orf132a EFKADVYVIGNVAKRGMDVVEAILNLGLFYISGPQWLSENVLHHHWLVGVAGTHGKTTT
    70 80 90 100 110 120
    70 80 90 100 110 120

35 orf132.pep SMLAWVLEYAGLAPGLIGGVXGKFR---RFRPFAANAAPRFEPQI-----AVFR
    orf132a SMLAWVLEYAGLAPGLIGGVXGKFR---RFRPFAANAAPRFEPQI-----AVFR
    130 140 150 160
    130 140 150 160

40 orf132.pep HRSRRIRHRLFRQTFFXIRALPSAYRVEQSGIRPRHLRCLGRDTPVPLPRAYRAXRL
    orf132a KRSKFVHYRPRTAVLNNLEFDHADIFADLGAITQFHHLVRTVPSEGLIVCNQGRQSLQD
    170 180 190 200 210 220
    180 190 200 210 220 230

```

45 The complete length ORF132a nucleotide sequence <SEQ ID 869> is:

```

1 ATGAACACACA TCCACATTAT CGGTATCGGC GGCACGTTTA TGGGTGGGAT
51 TGGCGCCATT GCCAAGAAG CAGSGTTTGA ATNCAGCGGT TCGATGCGCA
101 AGATGTATCC GCCATGTAC ACACACTCG AACGCTGCG CATAGCGTG
151 TATGAAGCT TCGACACCGC CAGTTTGAC GAATTTAAG CCGACGTTTA
50 201 CGTTATCGGC AATGTGCCA AGCGCGGAT GATGTGGT GAAGCGATT
251 TGAACCGTGG GCTGCTTAT ATTTCCGGCC CGCAATGGT GCGTGA AAC
301 NTGCTGCACC ATCATTTGNN ACTCGCGGT GCGGNGACG ACGGCAAAAC
351 GACCACCGCG TCTATGCTCG CTTGGGTTT GGAATATGCC GGACTCGCAC
55 401 CGGGCTTCNT TATCGCGGCG GTACCGGAAA ACTTCAGGT TCCGCCCCGC
451 CTGCGCGCAA CGCGCGGCGA AGACCGGAAC AGCCAAATCG CGTTTTTCGT
501 CATTTGAAGCC GACGAATAC ACACCGGTT TTTGCAAAA CGCTCCAAAT
551 TCGTGCAATTA CGTCCGCGT ACGCGCGGT TGAACAATCT GGAATTCGAC
601 CACGCGGACA TCTTGGCGGA TTTGGGCGCG ATACAGACCC AGTTCCACCA
551 CCTCGTGGT ACGTGGCGT CTGAAGGCT CATCGTCTGC AACGGACGCG
60 701 AGCAAGGCT CCAAGACACT TTGACAAAG GCTGCTGACG CGCGGTGGA
751 AATTTCGCA CGGAACACG CTGGCAGGCC GCGCAAGACA ATGCCGATGG

```

-470-

801 CTCGTTCGAC GTGTTGCTTG ACGGCAAAAA AGCCGGACAC GTCGCTTGGA
 851 GTTTGATGGG CGGACAACAC CGCATGAACG CGCTCGNGT CATCGCCGCC
 901 GCGCGCATG CGGAGTNGA CATTACAGAC GCGTCGGAAG CTTGAGCAC
 951 GTTTAAAAAC GTCAACCGCC GCATCGAAAT CAAAGCGACG GAAACCGTA
 1001 TCACCGTTTA CGACGACTTC GCCACCATC CGACCGGTAT CGAAACGACG
 1051 ATTCAGAGTT TGGCGCAGCG CTTCCGCGCG GCGCGCATC TCGCGCTGCT
 1101 CGAACCGCGT TCCAATACGA TGAAGCTGGG TACGATGAAA GCGCGCTGCG
 1151 CGCAAGCGT CAACAAGCC GACCAAGTGT TGTGTAAGC CGCGCGCTCG
 1201 GACTGGGACG TTGCGCAAGC CTTCCGCTGT TTGCGCGGCA GCGTCGACTG
 1251 OGCAGAACAG TTGCGTGCCT TCCTGCGCGA AATCGTGAAG AACGCGAAG
 1301 CAGCGGACCA TATTTTGGTG ATGAGCAACG GCGCTTTCGG CGGAATACAC
 1351 ACCAAACTCG TGGACGCTTT GAGATAG

This encodes a protein having amino acid sequence <SEQ ID 870>:

1 MKHIHIIGIG GTFMGGIAAI AKEAGFEXSG CDAKMYPPMS TQLEALGIGV
 51 YEGFDTAQLD EFKADVYVIG NVAKRGMDVV EAILNRGLPY ISGPQWLAEN
 101 XLHHHWXIGV AXTHGKTTTA SMLAWVLEYA GLAPGFIXIG VPENFSVSAR
 151 LPQTFRQDFN QSPPFFVIEA DEYDTAFDDK RSKFVHYRFR TAVLNNLEFD
 201 HADIFADLGA IQTFHHLVR TFPSEGLIVC NGRQCSLQDT LDKGCWTPVE
 251 KFTGHEGWQA GEANADGSFD VLLDGKAGH VAWSLMGHNN RMNALAVIAA
 301 ARHAGVDIQT ACEALSTFFN VKRRMEIKGT ANGIITYDD AHHPATAETT
 351 IQGLRQRVGG ARILAVLEPR SNTMKLGTAK AALPASLEKA QVFXIYAGGA
 401 DNDVAEALAP LGGRILHVGD FADFVAEIVK NAEAGSHLV MSNGGFGGIH
 451 TKLLDALR*

ORF132a and ORF132-1 show 93.9% identity in 458 aa overlap:

25 orf132a.pep MKHIHIIGIGGTFMGGIAAIAKEAGFEXSGCDKMYPPMSTQLEALGIGVYEGFDTAQLD
 orf132-1 MKHIHIIGIGGTFMGGIAAIAKEAGFEXSGCDKMYPPMSTQLEALGIDVYEGFDAQLD
 30 orf132a.pep EFKADVYVIGNVAKRGMDVVEAILNRGLPYISGPQWLAENXLHHHWXIGVAXTHGKTTTA
 orf132-1 EFKADVYVIGNVAKRGMDVVEAILNLGLFYISGPQWLENVLHHHWXIGVAGTHGKTTTA
 orf132a.pep SMLAWVLEYAGLAPGFIXIGVVPENFSVSARLPQTFRQDINSQSPPFFVIEADEYDTAFDDK
 35 orf132-1 SMLAWVLEYAGLAPGFIXIGVVPENFVSARLPQTFRQDINSQSPPFFVIEADEYDTAFDDK
 orf132a.pep RSKFVHYRFRPTAVLNNLEFDHADIFADLGAIQTFHHLVRTFPSEGLIVCNGRQCSLQDT
 orf132-1 RSKFVHYRFRPTAVLNNLEFDHADIFADLGAIQTFHYLVRTFPSEGLIVCNGRQCSLQDT
 40 orf132a.pep LDKGCWTFVEKFTGHEGWQAGEANADGSFVLLDGKAGHVAWSLMGGHNRNALAVIAA
 orf132-1 LDKGCWTFVEKFTGHEGWQAGEANADGSFVLLDGKTAGRVKWDLMGRHNRNALAVIAA
 45 orf132a.pep ARHAGVDIQTACEALSTFFKNVKKRMEIKGTANGITYDDFAHHPATAETTTIQLRQRVGG
 orf132-1 ARHAGVDIQTACEALSTFFKNVKKRMEIKGTANGITYDDFAHHPATAETTTIQLRQRVGG
 50 orf132a.pep ARILAVLEPRSNTMKLGTMKALPASLKEADQVFXIYAGGADNDVAEALAPLGRLHVGD
 orf132-1 ARILAVLEPRSNTMKLGTMKALPASLKEADQVFXIYAGGADNDVAEALAPLGRLHVGD
 orf132a.pep FADFVAEIVKNAEAGDHLVMSNGGFGGIHKLDAALRX
 55 orf132-1 FADFVAEIVKNAEAGDHLVMSNGGFGGIHKLDAALRX

Homology with a predicted ORF from *N.gonorrhoeae*

ORF132 shows 89.6% identity over 259 aa overlap with a predicted ORF (ORF132ng) from *N.gonorrhoeae*:

60 orf132.pep MKHIHIIGIGGTFMGGIAAIAKEAGFEXSGCDKMYPPMSTQLEALGIDVYEGFDAQLD 60
 orf132ng MKHIHIIGIGGTFMGGIAAIAKEAGFEXSGCDKMYPPMSTQLEALGIGVHEGFDAAQLE 60

	orf132.pep	EFKADYVYVIGNVAKRGMVVEAILNLGLFPIYSGPQWLSENVLHHHWLGVAGTHGKTTTA	120
	orf132.ng	 EFQADIYVIGNVARRGMVVEAILNLRLGFIYSGPQWLAEENVLHHHWLGVAGTHGKTTTA	120
5	orf132.pep	SMLAWLVEYAGLAPGLIGGVXGKFRFRFPFAANAAFRPEQPIAVFRHRSRRIHRHLRFQ	180
	orf132.ng	 SMLAWLVEYAGLAPGLIGGVPGKFRFRFPPTANAASRFEQQIAVFRHRSRRIHRHLRFQ	180
10	orf132.pep	TFXIRALPSAYRRVEQSGIRPRHLRLCRIGRDTDPVPLPRAYRAVKNLNLQRTAAKPARY	240
	orf132.ng	 TLQIRALSPAYRRVEQSGIRPRHLRLRIGRDTDPVPPFRAHRTIRRHRLQRTAAKPARY	240
15	orf132.pep	FGQRLLDAGGKIRHGTRLA	259
	orf132.ng	 FGQRLLDAGGKIRHRTLDW	261

1	MKHHIIIGIG	GTFFMGGLAI	AVARGKFKVS	CDAMKYPFMS	TQLEALGIGV
5	HFGDQAQLE	EFQADIYIV	NKARAGMDVV	EAILNRGLFV	LSGPQWALNV
10	VLHHHNVLGV	AGTHGKFTTS	SMI.AWVLEYA	GIAPGLFALG	VPKGFRFRFF
15	PRRAAASRPF	QQAIAVRTHS	RRIRHVFQRO	QTTLRISFA	YRVVQSGQIR
20	PRRLRLRLGR	DTDFVPFPFA	HRTIRRPRLR	QRTAAKFARY	FGQLDLAGG
25	KIRHTRLAD	W*			

25	1	ATGAAACATC	TCGCATATCT	CGGTATTCGG	GGACATCGAT	TGGGGCGGGAT
	51	TGCGCGCCATT	GCCAAARGAA	CCGGGTTCGA	AGATACGGGT	TGCACGACGAT
	101	AGATGTTATCC	GCGGATGATG	ACCACGCTCG	AAGCTCTGGG	CATAGGCGGT
	151	CACGAGGCGT	TCCGTGCGCG	CAGGTTGGAA	GAATTTCAAG	CGGATATTTA
	201	CGTCACTGGC	AATGTGCGCA	GGCGGGGAT	GGATGTGGTC	GAGGCGGATT
30	251	TGAAACGTGG	GTGCTGCTAT	ATTTTCGGGG	CGCAATGGGT	GCTCGAAAC
	301	GTGCTgcacc	atcatTTGGT	ATTCGGCGTG	gaggggaACG	ACGGcaAAC
	351	gaccaCcGcg	TCATCGTGTG	CTCGGTTGTT	GGAAATGCG	GMACTGCGC
	401	CGGGCTTCTC	ATCGCGGGGT	gtaccGGGAA	ATTTTCGGGT	TTCCGCGCG
	451	CTACGCGCAA	CGGCGGGTCA	gcagcAACAA	AGCAAAATCG	GTTTTTTTGT
35	501	CATCGARGAC	GAGCAATCAT	ACACCGCGTT	CTCGCAAAA	CGTCACTGAT
	551	TGCGTGGTGT	ACACGCGCTG	AGCGCGCTG	TGAACGCTG	AGTTCTCGAC
	601	CACGCGCACA	TCTTCGCGCA	CTTGGGCGGT	TATACAGACC	AGTTCCAAAC
	651	CTCGTGCGGC	ACCGTACCAT	CCGAGAGGCT	CATCGTGTG	AACGGACACG
	701	AGCAAAAGCT	GCAGATGATC	TATGGCAAG	GCTGTCTGAC	CGCGGTGGAA
40	751	AAATTTGCGCA	CGCGACACGG	CTGGCAGAT	GGTGAAGTCA	ATGCGCAGTG
	801	CTGCTGTGCT	GTAATGCTTT	ACGGCAAAA	AGCGGCAGT	CGCATATGGG
	851	ATTTGATGGG	CGGACACACA	CGCATGAAGC	CGCTCGCGCT	CATCGCTGCC
	901	GCACGCGCATG	CGCGAGTCGA	TGTTTCAAGC	GCGTGGGAAG	CTTGGGTGGT
	951	GTTTAAAAAC	GTCAAACCGC	GCATGGAAT	CAAAGGCGAC	CGAAACGGCA
45	1001	TCACCGTTTA	CGACGATTC	GCCCAACG	CGAACCGCAT	CGAAACAGCA
	1051	ATTTACAGGTT	TGCGCGCAAG	TGTCGCGGCT	CGCGGCGCAT	TGCGCGTCTC
	1101	CGAGCGCGGT	CCAAACACGA	TGAAACGTGT	CATCATATAG	TGCGGCTGCT
	1151	CGCGAGAGAG	CGCGGACGAT	GACCAAGTGT	CGCGGCGGCT	CGCGGCGGCT
	1201	GACTGGGACG	TGCGCGAAGC	CTCGCGGCT	TGGGCTGTCA	GCTCGCGGCT
	1251	CGGTAAGAAT	TTGCAACACT	CTGTTGCGCA	AATTTGTGAA	AGCGCGCGAA
	1301	CGCGGACACA	TATTTTGGTG	ATAGACAGAC	CGGGTTTCGG	CGGAATACAC
	1351	ACCAACCTCG	TGGACGCTTT	GAGATAG		

55		51	MKHIIHAIIGLE	FTPMGFGIAAI	AKEAGPKVSG	CDAKMYPFMS	TQLEALGIGVGV
		51	HGEFDAAGLE	EFQADIVYIE	NVARRGMDIV	EAILRLNGLE	ISQFOWLAENG
		101	VLLHHHVVGV	AGTGHGHTG	VWLVLELYA	GLAQGLQGT	VFENGVSYSAR
		151	LFPTFRDPN	DMTATVAFH	SVWVLEH	SVYLALENG	SVYLALENG
		201	HADIFATV	VTQTHHLE	TVSWLELIG	NGQQSLQD	LKDCWGVSR
		251	KFTGHGWOI	GEVNAADSG	VLLDGGKAG	VAMDLWNGH	RMMNALVIAAL
60		301	ARHAGVDVOT	ACEALGAFKN	VKRRMEIKGT	ANGTVDDYD	AHHFTAIETIT
		351	IQGLRVRVGG	ARILAVLRE	SNTHKLGTMK	NASPSALKEA	DOVFCYAGGA
		401	DMVDAALAP	LGCLRRLVLE	FUTVFAEIVK	HARTLGHVL	MSWGGFGGTH

ORF132ng-1 and ORF132-1 show 93.2% identity in 458 aa overlap:

	orf132ng-1.pep	NKHIHIIGIGGTFMGGIAAIAKEAGFKVSGCDAMYPMPSTQLEALGIVGHEGFDAQAQLE
5	orf132-1	NKHIHIIGIGGTFMGGIAAIAKEAGFEVSGCDAMYPMPSTQLEALGIVGHEGFDAQAQLE
	orf132ng-1.pep	EFQADIYVIGNVARRGMDVVEAILNRGLPYISGPQWLAENVLHHHVLGVAGTHGKTTTA
	orf132-1	EFKADYVIGNVAKRGMDVVEAILNRGLPYISGPQWLAENVLHHHVLGVAGTHGKTTTA
10	orf132ng-1.pep	SMLAWVLEYAGLAPGFLIGGVENFGVSARLPQTPRQDPNSKSPFFVIEADEYDTAFFDK
	orf132-1	SMLAWVLEYAGLAPGFLIGGVENFGVSARLPQTPRQDPNSKSPFFVIEADEYDTAFFDK
15	orf132ng-1.pep	RSKFVHYRPRPTAVLNNLEFDHADIFADLGAIQTFHHLVRTVPSEGLVNCNGQQQSLQDT
	orf132-1	RSKFVHYRPRPTAVLNNLEFDHADIFADLGAIQTFHHLVRTVPSEGLVNCNGQQQSLQDT
	orf132ng-1.pep	LDKGCWTPVEKFGTGHGWQIGEVNADGSDVLLDGGKAGHVAWDLGGHNRMLALAVIAA
20	orf132-1	LDKGCWTPVEKFGTGHGWQAGEANADGSDVLLDGGKTAGRWKDLGHRNRMLALAVIAA
	orf132ng-1.pep	ARHAGVDVQTACEALGAFKNVKKRMEIKGTANGITVYDDFAHHPTAIETTIIQGLRQRVGG
	orf132-1	ARHVGVDIQTACEALGAFKNVKKRMEIKGTANGITVYDDFAHHPTAIETTIIQGLRQRVGG
25	orf132ng-1.pep	ARILAVLEPRSNMKGTMKSLPASLKEADQVFCYAGGADWDVAEALPLGCRILRVGKD
	orf132-1	ARILAVLEPRSNMKGTMKSLPASLKEADQVFCYAGGVWDVAEALPLGGRILNVGKD
30	orf132ng-1.pep	FDTFVFAEIVKNARTGDHILVMSNGGFGGIHTKLLDALRX
	orf132-1	FDAPFAEIVKNAEVDGHIIVMSNGGFGGIHGKLLDALRX

In addition, ORF132ng-1 is homologous to a hypothetical *E. coli* protein:

35	pir S56459 hypothetical protein o457 - Escherichia coli >gi 537075 (U14003) ORF o457 [Escherichia coli] >gi 1790680 (AE000494) hypothetical 48.5 kD protein in fbp-pmba intergenic region [Escherichia coli] Length = 457 Score = 474 bits (1207), Expect = e-133 Identities = 249/439 (56%), Positives = 294/439 (66%), Gaps = 13/439 (2%)	
40	Query: 22	KEAGFKVSGCDAMYPMPSTQLEALGIVGHEGFDAQAQLEEFQADIYVIGNVARRGMDVVE 81 ++ G +V+G DA +YPPMST LE GI + +G+DA-QLE Q D+ +IGN RG VE Sbjct: 21 RQLGHEVTSGDANVYPPMSTLLEKQGIELIQGYDASQLEP-QPDLVIGNAMTRGNPCVE 79
45	Query: 82	AILNRGLPYISGPQWLAENVLHHHVLGVAGTHGKTTTASMLAWVLEYAGLAPGFLIGGV 141 A+L + +PY+SGPQWL + VL WVL VAGTHGKTTTA M W+LE G PGF+IGGV Sbjct: 80 AVLEKNIPYMSGPQWLHDFVLRDRWVLAAGTHGKTTTAGNATWILEQCQYKPGFVIGGV 139
50	Query: 142	PENFGVSARLPQTPRQDPNSKSPFFVIEADEYDTAFFDKRSKFVHYRPRPTAVLNNLEFDH 201 P NF VSA L +S FFVIEADEYD AFFDKRSKFVHY PRY T+LNNLEFDH Sbjct: 140 PGNFVESAHL-----GESDFVIEADEYDCAFFDKRSKFVHYCPRTLILNNLEFDH 190
55	Query: 202	ADIFADLGAIQTFHHLVRTVPSEGLVNCNGQQQSLQDTLDKGCWTPVEKFGTGHGWQIG 261 ADIF DL AIQ QFHHVL VP +S I+ GCV+ E S WQ Sbjct: 191 ADIFDDLGAIQTFHHLVRIVPGQGRITWPEINDILKQTMAMGCWSEQLVEQGHWQAK 250
	Query: 262	EVNADGSDVLLDGGKAGHVAWDLGGHNRMLALAVIAAARHAGVDVQTACEALGAFKN 320 ++ D S ++VLLDG+K G V W L+G HN N L IAAARH GV A ALG+GFN Sbjct: 251 KLPTDASEWEVLLDGEKGVGVKSLVGEHNMHNLATAAARHVGVPADAAANALGSTN 310
60	Query: 321	VKKRMEIKGTANGITVYDDFAHHPTAIETTIIQGLRQRVGG-ARILAVLEPRSNMKGTM 379 +RR+E++G ANG+TVYDDFAHHPTAI T+ LR +VGG ARI+AVLEPRSNMKG+G Sbjct: 311 ARRLLELRGEANGTVYDDFAHHPTAILATLALRGKVGSTARI+AVLEPRSNMKGKIC 370
65	Query: 380	KSLPASLKEADQVF-CYAGGADWDVAEALPLGCRILRVGKDFDTFVFAEIVKNARTGDHIL 438 K L SL AD+VF D T W +V+E GHIL Sbjct: 371 KDDLAPSLGRADEVFLQPAHIPQWAEVAEACVQPAHWSGDVDTLADMVKTQAQGHIL 430
	Query: 439	LVMSNGGFGGIHTKLLDAL 457

LVMSNGGFGGIIH KLLD L
 SbJct: 431 LVMSNGGFGGIIHQLDGL 449

Based on this analysis, it was predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

- 5 ORF132-1 (26.4kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 20A shows the results of affinity purification of the His-fusion protein, and Figure 20B shows the results of expression of the GST-fusion in *E.coli*. Purified His-fusion protein was used to immunise mice, whose sera were used for FACS analysis (Figure 20C) and ELISA (positive result). These
 10 experiments confirm that ORF132 is a surface-exposed protein, and that it is a useful immunogen.

Example 103

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 875>

```

1  ..CGCGGCGTATT ACGGCTCGGA TGACGAATTT AAGCGGGCAT TCGGAGRAAA
51  CTCGCGCGACA TmCAAGAAAC AITGCAACCG GAGCTGCGGG ATTATGAAC
15  CGGTATTGAA AAATAACGGC AAAAAGCGCG CCAACACCA TCGGTGAGC
151  ATTAGTGGCG ACTTCGGCGA TTAITTCATG CGGTGCGCA GATATTGCG
201  CACACACCGT ATGCCACACA TCAAGAAAT GTATTETTC CAATTCGGCG
251  ACTCGGCGST TCACACCGCC TTAACAACAG ACGGCGCAA CATTGCGCAA
301  TTTGGCTTCr ATACCTATAA AAAAGGATTG TTAACAACAG ATGATACATT
351  AGGATTAAAA CTGGTCGGCT ACGCGACGCG CATCGACAC TACATCCACA
401  ACGTTTACGG GAATGTGGG GATTTGAAAC GGGATATTCC GAGCTGGGTG
451  AGCAGCACCG GCCTTGCGCTA CACCATCCAA CATCGCrATT TCAWAGACAA
501  AGTGATCAAA nnnnnnnnnn nnnnnnnnnn nnnnTACGAT TATGGGCGTT
551  TTTTCCACCA CTTTCTTAC GCCTATCAAA AAAGCACGCA ACCGACCACT
25  TTCAGCGATG CGAGCGAATC GCCCAACAA CGCTCCAAAG AAGACCAACT
601  CAAACAAAGT TATGGGTTGA GACGGGTTTC CGCCTCGCG CGAGATTACG
651  GACGTTTGGa AGTCGGTACG CGCTGGTTGG GCAACAAACT GACTTTGGCG
701  GGCGCGATGC GCTATTTCGG CAAGACATFC CGCGCGAGCG CTGAGACACG
751  CTATATCGAC GGCACCAACG GGGGAATAC CAGCAATTTC CGGCACTCG
801  GCAGCGTTC CATCAACAA ACCGAATCTC TTGCGCCGCA GCCTTTGATT
30  TTWgATTTTa ACGCGCGTTA CGAGCGGAAG AAAAACCTTA TTTCCGCGCG
901  CGAAGTCAAA AATCTGTGCG ACAGCGGTTA TATCGATCG CTGATCGCG
951  GCAATGATGC GGCACr.GAG CGTATTATTA GCTCGTTGCA CCGCAAGAC
1001  AAGGACrAG ACGTAAACGT TAATGCTGAT AAAACGTTGT GcaACGGCAA
1051  ATACGCGCGC ACAGCAAAAG CGGTATTGAC CAATTTTCCA CGCGGACGCA
1101  CTTTTTTgAT GACGATGAGC TACAAGTTTT AA
1151

```

This corresponds to the amino acid sequence <SEQ ID 876; ORF133>:

```

1  ..PGYYGSDDFF KRAFGENSPT KKKHCNRSOG IYEPVLKRYG KKRANHHVS
51  ISADFGDYFM PFASYSRTHR MPNIQEMYFS QIGDSGVHTA LKPERANTWQ
40  PGFXTYKGLK LKQDDTLGLK LVGYSRSDIN YIHNVYGRKW DLNGLDPSWV
101  SSGTGLAYTIQ HRXFDKRVHQ XXXXXXXXXD YGRFFTNLSY AYQKSTQPTN
151  FSDASESPNN ASKEDQLKQG YGLSRVSLP RYGRLELVGT RMLGNKLTG
201  GAMRYFGKSI RATAERYID CTNGGNTSNF RQLKRSIKQ TETLARQLI
251  XDFNAAYEPK KNLIFRAEVK NLEFRYIDP LDAGNDAAKE RYYSFDFPKD
45  KDXDVTNCAD KILGSKYGG TSKSVLNFRA RGRTELMTMS YK*
301

```

Further work revealed the further partial DNA sequence <SEQ ID 877>:

```

1  GAGGCGCAGA TACAGGTTTT GGAAGATGTG CACGTCAAGG CGAAGCGCGT
51  ACCGAAGACG AAAAAGTGT TTACCGATGC GCGTCCGCTA TCGACCGCTC
101  AGGATATATT CAATCCAGC GAAACCTCG ACAACATCTG ACGCAGCATC
151  CCGCGTGGT TTACACAGCA AGATAAAGC TCGGCGATTG TGTCTTTGAA
201  TATTCGCGCG GACAGCGGGT TCGGCGGGT CAATACGATG GTGACGCGCA
251  TCACGACAGC CTTTATTCTG ACTTCTACCG ATCGGCGGAG GCGAGCGGGT

```

-474-

5	301	TCAATCTCAAT	TCGGTGCATC	TGTGCACAGC	AAITTTTATTG	CCGGACTGGA
	351	TGTCGTCAAA	GGCAGCTTCA	GCGGCTCGGC	AGGCATCAAC	AGCCTTGCCG
	401	GTTCGGGAA	TCTCGCGACT	TTAGCGCTGG	ATGACGTCGT	TACGGGCAAT
	451	AATACCTACG	GCTCTGTGCT	AAAGGTCCTG	ACCGGCACCA	AITTCAACCAA
	501	AGGTAAATGG	ATGGCGGGGA	TAGGTGCGCG	CAAAATGGCTG	GAAAGCGGAG
	551	CATCTGTGCG	TGTGCTTTAC	GGGCACAGCA	GGCGGAGCTG	GGCGCAAAAT
	601	TACCGGTGGG	GCGGCGCGCG	GCAGCACATC	GGAATTTTGG	GCGCGGAAAT
	651	TTTGAGTATTC	CTATTGGGCG	CGATTTTAAAG	GGATTAAGAT	CGTTTGAATC
	701	TCAAATTCGGA	CAGCGGAAAA	TGGGACGCGG	ATTTTCAAAAG	GCAACACTGG
10	751	AAATCAACAG	CCTATAAAAA	TTACAACAAC	CAAGAATCTAC	AaAAATACAT
	801	CGAAGAGCAT	GACMAAAGCT	GGCGGAAAA	CCTg, CaCG	CAATACGACA
	851	TTACCCCATC	CGATCGCTCC	AGCCTGAGCG	AGCAGTCGCG	AGGCAATCTG
	901	TTTAAATTTGG	AATACGACGG	CGTATTCAAT	AAATACAAGC	CGCAATTTCC
	951	CGATTTAAAC	ACCAAAATCG	GCAGCGCGAA	AATCATCAAC	CGCAATTATC
15	1001	AGTTCAATTA	CGGTTTGTCT	TTGAACCGGT	ATACCAACCT	CAATCTGACC
	1051	GCAGCTACCA	ATTCGGGCGAG	GCAGAAATAT	CGAAAGGGGT	CGAAGTTTAC
	1101	AGGCTGGGGG	CTTTTAAAGG	ATTTTGAAC	CTACAACAAC	GGGAAATCC
	1151	TCGACCTCAA	CAACACCGCC	ACCTTCGCGG	TGCCCGCGGA	AACCGAGTTG
20	1201	CAAAACCATT	TGGGCTTCAA	TATTTTCCAC	AAGCAATACG	GCAAAAACCG
	1251	CTTTCTCTAA	GAATTTGGGG	TGTTTTTCSA	CGGTCTGCTAG	CAGGACAAAG
	1301	CGCTTTATTC	CTATTGGGCG	CGATTTTAAAG	GGATTAAGAT	CGTTCTGCGC
	1351	CAAAATCAAA	CGATTTGTCCA	ACCGCGCGCG	AGCCAAATAT	TCAACAGCTT
	1401	CTACTTCGAT	GCGCGCTCCA	AAAAAGACAT	TTACGCTTAA	AACTACAGCA
25	1451	CCAAATACCT	CGGCTACCGT	TTCCGCGCGG	AAATATACGG	CTATTACCGG
	1501	TCGGATGAGC	AATTTAAGCG	GGCATTCCGA	GAAATCTCGC	CGACATACAA
	1551	GAAACATTTGC	AACCGGAGCT	GCGGGAATTA	TGAACCCGTA	TTGAAAAAAT
	1601	ACGGCAAAAA	GCGCGCGAAC	AACCAITTCG	TGACGATTAG	TGCGGACTTC
	1651	GGCGATTATT	TCAATGCCCTT	CGCCAGCTAT	TGCGGCAACG	ACCGTATGCC
	1701	CAACATCCAA	GAAATGTATT	TTTCCCAAT	GGCGGACTCC	GGGCTTCAAC
30	1751	CGCGCTTAAA	ACCAGACGCG	GCAAAACACT	GGCAATTTGG	CTTCAATACC
	1801	TATAAAAAAG	GATTTGTAAA	ACAAGATGAT	ACATTAGGAT	TAAATCTGGT
	1851	CGGCTACCGC	AGCGGCAATC	ACAACATACAT	CCACAACCTT	TACGGGAAT
	1901	GCTGGGATAT	GACGGGGAT	ATTTCCAGCT	GGGTGAGGAG	CACCGGGCTT
	1951	CGCAACGCTA	TCACATACCG	CAATTTTCAA	GCAAAATATG	ACCAAAATATG
35	2001	TTTGTGATTG	GAGCTGAAT	ACGATATATG	CGCTTTTTC	ACCAATCTTT
	2051	CTTACGCTA	TCAAAAAAGC	AGCAACCGGA	CCAACCTTCAG	CGATGAGGAC
	2101	GAATCGCCCA	ACAATGCGCT	CAAGAAGAC	CAACTCAAC	AAGGTTATGG
	2151	GTGAGCAGG	GTTCGCGCC	TGCGCGGAGA	TTACGGAGCT	TTGGAAGTCG
40	2201	GTACGCGCTG	GTTCGGCAAC	AACTGACTTT	TGGGCGCGCG	GATGCGCTAT
	2251	TTGGCGAAGA	GCATCGCGCG	GACGGCTGAA	GAACGCTATA	TCGACGGCAC
	2301	CAACGGGGGA	AATACAGGCA	ATTTCCGSCA	ACTGGGCAAG	CGTTCCATCA
	2351	AAACAAACCGA	AACCTGTGCC	CGCCAGCCTT	TGATTTTGA	TTTTTACGCC
	2401	GCTTACGAGC	CGAAGAAAAA	CCTATTTTTC	CGCGCGGAAG	TCAAAATCT
45	2451	GTTCGACAGG	CGTTATATCG	ATCCGCTCGA	TGCGGGCAAT	GATGCGGCAA
	2501	CGCAGCGTTA	TTACAGCTCG	TTGACCGCGA	AAGACAGGCA	CGAAGACGTA
	2551	ACGCTGATAT	CTGATAAAC	TGTTGCAAC	GGCAATACG	CGGCGACGAG
	2601	CAAAAGCTA	TGACCAATT	TTGACGCGGG	ACGCACTTT	TTGATGACGA
	2651	TGAGCTACAA	GTTTTAA			

This corresponds to the amino acid sequence <SEQ ID 878; ORF133-1>:

50	1	EAQIQVLEDV	HVKAKRVFKD	KKVFTDARAV	STRQDIFKSS	ENLNDIVRSI
	51	PGAGTQDKKS	SGIVSLNIRG	DSGFGRVNTM	VDGITQTFFS	STDAGRAGG
	101	SSQPGASVDS	NFIAGLDVVK	GSFSGSAGIN	SLAGSANLRT	LGVDVVQGN
	151	NTYGLLLKLG	TGNTSTKGNA	MAIAGARKWL	ESGASGVGLY	GHSRRSVAQN
	201	YRVGGGGQHI	GNPGEAYLER	RKQRYFVQEG	ALKFNSDSGK	WERDLQRQW
55	251	KYKPYKNYNN	QELQKYIEEH	DKSWRENLPX	QYDITEIDPS	SLKQGSAGNL
	301	FKLEYDGVFN	KYTAQFPRDLN	TKIGSRKIIN	RNYQFNYGLS	LNPTYNLNL
	351	AAYNSSGRQKY	PKGSKFTGWL	LLKDFETYNN	AKILDLNNTA	TFRLEPRETEL
	401	QTTLGFNYPH	NEYGKNRFPF	ELGLFFDGD	QDNGLSYSLG	RFKGDKGLLF
60	451	QKSTIVQPAH	SGYFNFTFFD	AALKKDIYRL	NKSTNVCYGR	RGGEYTGYYG
	501	SDDEPKRAFG	ENSPETKHKC	NSGCGIHEVP	LKYGKGRAN	NHVSLSIADP
	551	GDYEMFASV	SRTHHEMLQ	EMVFSQIGDS	GWHALKPERP	ANWYQPGNT
	601	YKGLGLKQDD	PLGLKLGVRY	SRIDNYHNHV	YKGWMDLND	IPSWVSTGL
	651	AYTIQHRNEK	DKVHKHGFEL	ELNDYDGRFF	TNLSYAYQKS	TQPTNFSDA
	701	ESPNASKEDK	QLKQGYGLSR	VSALEPRDYG	LEVGTWRLGN	KITLGGAMRY
65	751	FGKSIRATAE	ERYIDGTNGG	NTSNFRQLGK	RSIKQTEYSL	RQLPLIFDFYA
	801	AYEPKKNLIF	RAEVKNLFDR	RYIDPLDAGN	DAATQRYTSS	FDPKDKDEVD
	851	TGNADTKLCN	GRYGGTSKSV	LNTNFRGRTF	LMTMSYKF*	

Computer analysis of this amino acid sequence gave the following results:

[illegible]

[illegible]

51	AAGACAGAA	ANGTGTTCAC	CGATGTGGTTC	CGCGTATGAC	CCCGTCGACG
1	TATTTTCAAA	TCGACGAGAA	ACCTCGACGAA	CATCTGACGC	ACCGTCACCG
101	GTGCGTTCAC	ACGACGAGAA	AAAAGCTCGG	CNNTGTGGT	TTTGAATATC
151	CGCGGACGAA	CGCGGTTCCG	GCGGTTGATG	ACNATGTGTC	ACGGCATCAC
201	NCANACCTTT	TATTTGACCT	TACCGATGTC	GCGGAGCGCA	CGCGGTCATC
251	CTCAATGCTG	TGCTATCTGC	GACAGACATG	TTATNCGTCG	ACTCGATGAC
301	CTGAAAGCGA	CTCTACGGCG	CTGGGACGAG	ATCAACAGCC	TTGCGGTGCT
351	GCGCATATCT	GCGGCTTTAN	GCTGGGATGA	TGTGCTTCAG	GCGCATATTA
401	CNATCGGCTC	CGCGTAAAG	CTGCTGACGC	GACCATATTC	CCAAAGAGAT
451	AATGCACTGC	GCGGATGATG	CTGCTGAGAG	CTGCTGAGAG	CGACGATCAT
501	CGGCGGAGCG	CGGCGGAGCG	CGACAGCGAG	CACCTGCTTC	CAAAATATGAC
551	GCGTGGGCGC	CGCGGGACAC	CACATCGGAA	TTTGTGGCGC	GCGAATATCT
601	GAGACGACGA	ACGACATGCG	TTTGTGACGA	GAGGCGGCGT	TGAAATTCGAA
651	TTTCCACAGC	CGAAAATGCG	AGCGGGATTT	CGAAAGTGTG	TACTCGAAAA
701	CGAGTGGGTA	TCAAAATATG	AGCGCCCC	AGAGCATGCA	AAATATGACG
751	GAGGTCGTAT	ATTAAGAGAG	CGCGGAAAGC	CTGGCGGCGC	AGATACGCAT
801	CACCCCATCT	GATCGTGCCA	GCTGAAGCN	GAGTGTGCGA	GGGACACTGT
851	TTAAATATGC	ATGACGACGC	GTAATTCAAT	AATACGACGC	GCAATTTATG
901	GATTTAAACA	CGAAATATGC	CAGCGCGGCA	ATCTGACGAC	CGAATTTATG
951	CGGCGGAGCG	TTTGTGAGAG	CGGCGGAGCG	CGGCGGAGCG	CGGCGGAGCG
1001	AGGCTACATA	CGGCGGAGCG	CAGAAATATG	CGAAGGCTAT	GAGTTTACAT
1051	GCGTGGGGCG	TTTNTAAGAG	TTTNTAAGC	TACACAAAGC	CAAAAATCTC
1101	CGACCTCAGC	ACACACTCTC	CTCTCGGCT	CCCGCGTGAA	ACCGGATTCG
1151	AAACCATCTT	GGGCTTCAAT	TATTTTCAAC	ACGAAATGCG	CAAAAATGCG
1201	TTTCTGGAAG	TATTTGGGCT	GTITTTTCCG	GTTCGCGAAT	ANGACACAGC
1251	GCTTTATPTC	TATTTGGGCG	GTTTNTAAGG	TGCAATAGG	CTGCTGCCCC
1301	AAAAATCAAT	CATCTTGCCA	ACCGCGCGCA	GCGAATATTC	CACACGCTGT
1351	TACTTCGATG	CGCGGCTCAA	AAAAGACAT	TACCGCTTAA	ACTACGACGAC
1401	CATTCACGCT	CTGCTACGCT	TCGCGCGGNA	ATATACGGGC	TATTTACGCT
1451	CGGCGGAGCG	ATTTGAGGAG	CGGCGGAGCG	CGGCGGAGCG	CGGCGGAGCG
1501	AAGATGATGA	CGGAGCTATG	CGGATATGAG	GACCGGCTAT	TGAAAATAAT
1551	CGCGAAAAG	CTGCGGCAACA	ACCAATCGGT	CAGCATTAAT	CGGGACTTGC
1601	GCGATATATC	TATCGGCTGT	CGCAGTATGT	CGGCGACAGA	CGGTATGCGC
1651	ACCATTCAGG	AAATGTATG	TTTCCCAATG	GGGACTCGCT	GCGTTCACAC
1701	CGCTTAAAGC	CACAGCGCGG	CAACACATTC	GCAATTTGGC	TTTCAATACAT
1751	ATAAAAAGAG	ATTGTATAAA	CAGATGATAT	TATTAGGAT	AAATCTGTGT
1801	GCGTACGCGA	GCGCGATGCA	CNACATGATC	CACACAGTTT	CGCGGAAATG
1851	TGGGGATTTG	AGCGAGGAGC	TTTCCAGGCT	GTTGACGAGT	ACCGGGCTGT
1901	CTCACACCAT	CCACACCACT	AATTTCAABG	CAAAAGTGCA	CGACACCGGT
1951	TTTGTAGGTC	AGCTGATATG	CGATGATGAG	GTTTGTATCT	CGACACCGGT
2001	CGGCGGAGCG	CGGCGGAGCG	CGGCGGAGCG	GATGCGCGAG	GATGCGCGAG
2051	AATCGGCCAA	CAATGTGCTC	AAAGAGAGCG	AACTCAACAA	AGGTTATGGG
2101	TTGACGAGGG	TTTTCGCCCT	CGCCCGAGAT	TACGAGCATG	TGGAAGATCTG
2151	TACGGCGTGT	TGGGCAACGA	ACTGACATTT	GGGCGGGGCG	ATCGCGTATG
2201	TGCGCAAGAG	CATCGCGGCG	ACGCGTGAAG	AGCGCTATAT	CGACGCGGAG
2251	AATGGGAGGN	NTACACGAA	CTGCGGCAAT	CTGGGCAATG	GTTGCCATCA
2301	ACAAACGCGA	ACCTCTGCCAT	CGACGCTGTT	GATTTTGAAT	TTNTACATCG
2351	CTTACGAGCG	GAGAGAAAAA	CTTTTFTTC	GCGCGCAAGT	CAAAAATCTG
2401	TTGCAGACAG	GTTATATGTC	TCGCGTCAAT	GCGGSCAATG	ATGCGGACGAG
2451	CGAGCGTTAT	TACAGTTGCT	TCGACGCGAG	AGACGACATG	GACGAGATAT
2501	CGGCGGAGCG	CGGCGGAGCG	CGGCGGAGCG	CGGCGGAGCG	CGGCGGAGCG
2551	AAAGGTGTAT	TGCAATTAAT	TGCAAGATAT	CNACATCTT	TGATAAGATAT
2601	GAGCTACAGT	TTTTTA			

This encodes a protein having (partial) amino acid sequence <SEQ ID 880>:

```

1  KDKKVFETDAR AVSTRQDIFK SXENLDNIVR XIPGFTXQX KSSGXVSLNI
51 RXDSGFGGRVN TMDVGITXTF YSTSTDAAGRA GSSSQFGASV DSNFXAGLDV
101 VKGSFSGSAG INSLGASANL RTLXVDVVQ GNXYTGLLLK GLTGTNSTKG
151 NAMAATIGARK WLESGASGVV LYGHSRRSVA QNYRVGGGGQ HIGNFGAEYL
201 ERRKQRYFEQ EGGKLFNSNS GKWERDFQKS YWTKWYQKY DAPQLGKYI
251 EGHDKSWREN LAPQYDITPI DPSSLKQSA GNLFKLEYDG VFNKYTAQFR
301 DLNLTIGSRK IINRNQYFN YGLSLNPTNL NLTAAYNSGR QKYPKGSKPT
351 GWGLXKDFET YNNAKILDLL NTSTFRLPRE TELQTLTGFN YFHNEYGKNR
401 FPEELGLFFD GPDXDNGLYS YLGRFKGDKG LLPQKSTIVQ PAGSQYFNTF
451 YFDAALKKDI YRLNYSNTV YGRFGKGYTG YYXSDDEFRK AFGENSPTYX
501 KHCQSCGIIY EPLRKYGKK RANHSVSIS ADFGDYFMPF ASYSRTHRMP
551 NIQEMFYFSQI GDSGVHTALK PERANTWQFG FNTYKGLLK QDDILGLGLV
601 GYRSRDIKXI HNVYQKWWL NGRIPSWVSS TGLAYTTQHR NFKDKVHKH
651 FELELNYDYX RFTNLSLAY QKSTOPTNFS DASESPNNA KEDQLQCYG
701 LSRVSALPRD YGRLEVGTWR LGMKILTGA MRYFGKSIRA TAEERYIDXT
751 NGXXTSNFRQ LGKRSIXQTE TLAQPLIFD XYAAEYKPKX LIFRAEVKNL
801 FDRRYIDPLD AGNDAATQRY YSFDPKDKD EEVTCNDNNT LCNKGKYGTS
851 KSVLTNFARG XTFLITMSYK F*

```

20 ORF133a and ORF133-1 show 94.3% identity in 871 aa overlap:

```

                                     10      20      30      40
orf133a.pep      KDKKVFETDARAVSTRQDIFKSSXENLDNIVRXIPGFTXQXKX
25 orf133-1      EAQIQVLEVDVHKAKRVFKDKKVFETDARAVSTRQDIFKSSXENLDNIVRSIPGFTXQXKX
                                     10      20      30      40      50      60
orf133a.pep      SGXVSLNIRXDSGFGGRVNTMDVGITXTFYSTSTDAAGRGSSQFGASVDSNFXAGLDVVK
30 orf133-1      SGIVSLNIRGDSGFGGRVNTMDVGITQTFYSTSTDAAGRGSSQFGASVDSNFXAGLDVVK
                                     70      80      90      100     110     120
orf133a.pep      GSFSGSAGINSLGASANLRTLXVDVVQGNXYTGLLLKGLTGTNSTKGNAMAATIGARKWL
35 orf133-1      GSFSGSAGINSLGASANLRTLXVDVVQGNNTYGLLLKGLTGTNSTKGNAMAATIGARKWL
                                     110     120     130     140     150     160
orf133a.pep      GSFSGSAGINSLGASANLRTLXVDVVQGNXYTGLLLKGLTGTNSTKGNAMAATIGARKWL
35 orf133-1      GSFSGSAGINSLGASANLRTLXVDVVQGNNTYGLLLKGLTGTNSTKGNAMAATIGARKWL
                                     130     140     150     160     170     180
orf133a.pep      ESGASVGVLYGHSRRSVAQNYRVGGGGQHGNFGAEYLERRKQRYFEQEGGLKFNNSGK
40 orf133-1      ESGASVGVLYGHSRRSVAQNYRVGGGGQHGNFGAEYLERRKQRYFVQEGALKFNNSGK
                                     170     180     190     200     210     220
orf133a.pep      ESGASVGVLYGHSRRSVAQNYRVGGGGQHGNFGAEYLERRKQRYFVQEGALKFNNSGK
40 orf133-1      ESGASVGVLYGHSRRSVAQNYRVGGGGQHGNFGAEYLERRKQRYFVQEGALKFNNSGK
                                     190     200     210     220     230     240
orf133a.pep      WERDFQKSYWKTKWYQKYDAPQLQKYIEGHDKSWRENLAPOYDITPIDPSSLKQQSAGN
45 orf133-1      WERDLRQCKQYKFNKYNNTN-QELQKYIEGHDKSWRENLAPOYDITPIDPSSLKQQSAGN
                                     230     240     250     260     270     280
orf133a.pep      WERDFQKSYWKTKWYQKYDAPQLQKYIEGHDKSWRENLAPOYDITPIDPSSLKQQSAGN
45 orf133-1      WERDLRQCKQYKFNKYNNTN-QELQKYIEGHDKSWRENLAPOYDITPIDPSSLKQQSAGN
                                     250     260     270     280     290
orf133a.pep      LFKLEYDGVFNKYTAQFRDLNLTIGSRKIINRNQYFN YGLSLNPTNLNLTAAYNSGRK
50 orf133-1      LFKLEYDGVFNKYTAQFRDLNLTIGSRKIINRNQYFN YGLSLNPTNLNLTAAYNSGRK
                                     290     300     310     320     330     340
orf133a.pep      LFKLEYDGVFNKYTAQFRDLNLTIGSRKIINRNQYFN YGLSLNPTNLNLTAAYNSGRK
50 orf133-1      LFKLEYDGVFNKYTAQFRDLNLTIGSRKIINRNQYFN YGLSLNPTNLNLTAAYNSGRK
                                     300     310     320     330     340     350
orf133a.pep      YPKGSKFTGWGLXKDFETYNNAKILDLNNTATFRLPRETELQTLTGFNYFHNEYGKNRFP
60 orf133-1      YPKGSKFTGWGLXKDFETYNNAKILDLNNTATFRLPRETELQTLTGFNYFHNEYGKNRFP
                                     350     360     370     380     390     400
orf133a.pep      YPKGSKFTGWGLXKDFETYNNAKILDLNNTATFRLPRETELQTLTGFNYFHNEYGKNRFP
60 orf133-1      YPKGSKFTGWGLXKDFETYNNAKILDLNNTATFRLPRETELQTLTGFNYFHNEYGKNRFP
                                     360     370     380     390     400     410
orf133a.pep      EELGLFFDGPDXDNGLYSYLGRFKGDKGLLPQKSTIVQPAGSQYFNTFYFDAALKKDIYR
65 orf133-1      EELGLFFDGPDXDNGLYSYLGRFKGDKGLLPQKSTIVQPAGSQYFNTFYFDAALKKDIYR
                                     410     420     430     440     450     460
orf133a.pep      EELGLFFDGPDXDNGLYSYLGRFKGDKGLLPQKSTIVQPAGSQYFNTFYFDAALKKDIYR
65 orf133-1      EELGLFFDGPDXDNGLYSYLGRFKGDKGLLPQKSTIVQPAGSQYFNTFYFDAALKKDIYR
                                     420     430     440     450     460     470

```

-478-

		470	480	490	500	510	520
	orf133a.pep	LNYSNTNVGYRFGGXYTGYYXSDDEFKRAFGENSPYTKKHCHNRCSCGIEYEPVLKRYGKKRA					
5	orf133-1	LNYSNTNVGYRFGGXYTGYYXSDDEFKRAFGENSPYTKKHCHNRCSCGIEYEPVLKRYGKKRA	480	490	500	510	520
	orf133a.pep	NNHSVSIADFGDYFMPFASYSRTHRMPIQEMYFSQIGDSGVHTALKEPERANTWQPGFN					
10	orf133-1	NNHSVSIADFGDYFMPFASYSRTHRMPIQEMYFSQIGDSGVHTALKEPERANTWQPGFN	540	550	560	570	580
	orf133a.pep	TYKGLLKQDDILGLKLVGYRSRIDXYIHNHYGKWWDLNGNIPSWVSTGLAYTIQHRNF					
15	orf133-1	TYKGLLKQDDILGLKLVGYRSRIDXYIHNHYGKWWDLNGNIPSWVSTGLAYTIQHRNF	600	610	620	630	640
	orf133a.pep	KDKVKHKGFELELNVDYGRFPTNLSYAYQKSTQPTNFSDAESPNNASKEDQLKQGYGLS					
20	orf133-1	KDKVKHKGFELELNVDYGRFPTNLSYAYQKSTQPTNFSDAESPNNASKEDQLKQGYGLS	660	670	680	690	700
	orf133a.pep	RVSALPRDYGRLEVGTWRWLGKLTGGLAMRYFGKSIIRATAEERYIDXTNGXXTNSFRQLG					
25	orf133-1	RVSALPRDYGRLEVGTWRWLGKLTGGLAMRYFGKSIIRATAEERYIDXTNGXXTNSFRQLG	720	730	740	750	760
	orf133a.pep	KRSIKQTETLARQPLIFDXYAAYEPKKNLIFRAEVKNLFDRIYIDPLDAGNDAATQRYYS					
30	orf133-1	KRSIKQTETLARQPLIFDXYAAYEPKKNLIFRAEVKNLFDRIYIDPLDAGNDAATQRYYS	780	790	800	810	820
	orf133a.pep	SFDPKDKDEEVTCTNDNTLCNGKYGGTSKSVLTNFARGXTFLITMSYKFX					
35	orf133-1	SFDPKDKDEEVTCTNDNTLCNGKYGGTSKSVLTNFARGXTFLITMSYKFX	840	850	860	870	880
	orf133a.pep	SKEDQLKQGYGLSRVSALPRDYGRLEVGTWRWLGKLTGGLAMRYFGKSIIRATAEERYIDG					
40	orf133-1	SKEDQLKQGYGLSRVSALPRDYGRLEVGTWRWLGKLTGGLAMRYFGKSIIRATAEERYIDG	840	850	860	870	880

Homology with a predicted ORF from *N. gonorrhoeae*ORF133 shows 92.3% identity over 392 aa overlap with a predicted ORF (ORF133ng) from *N.*45 *gonorrhoeae*:

	orf133.pep	PGYYGSDDEFKRAFGENSPYTKKHCHNRCSCGIE	31
	orf133ng	FYFDAALKKDIYRLNYSNTNINRYRFGGXYTGYYXSENEFKRAFGENSPYKHECDPSCGL	560
50	orf133.pep	YEPVLKRYGKKRANNNHSVSIADFGDYFMPFASYSRTHRMPIQEMYFSQIGDSGVHTAL	91
	orf133ng	YEPVLKRYGKKRANNNHSVSIADFGDYFMPFAGYSRTHRMPIQEMYFSQIGDSGVHTAL	620
55	orf133.pep	KPERANTWQPGFXYTKKGLKQDDILGLKLVGYRSRIDXYIHNHYGKWWDLNGNIPSWV	151
	orf133ng	KPERANTWQPGFNYYTKKGLKQDDILGLKLVGYRSRIDXYIHNHYGKWWDLNGNIPSWV	680
	orf133.pep	STGLAYTIQHRNFKDKVHQXXXXXXXXXXDYGRFPTNLSYAYQKSTQPTNFSDAESPNN	211
60	orf133ng	STGLAYTIHRNFKDKVHQXXXXXXXXXXDYGRFPTNLSYAYQKSTQPTNFSDAESPNN	740
	orf133.pep	SKEDQLKQGYGLSRVSALPRDYGRLEVGTWRWLGKLTGGLAMRYFGKSIIRATAEERYIDG	271
	orf133ng	SKEDQLKQGYGLSRVSALPRDYGRLEVGTWRWLGKLTGGLAMRYFGKSIIRATAEERYIDG	800
65	orf133.pep	TNGNTSNFRQLGKRSIKQTETLARQPLIXDFAAYEPKKNLIFRAEVKNLFDRIYIDPL	331
	orf133ng	TNGNTSNVRQLGKRSIKQTETLARQPLIFDXYAAYEPKKNLIFRAEVKNLFDRIYIDPL	860

```

orf133.pep  DAGNDAAXERYSSFDPKDKDXDVTGNADKTLNCGRYGGTSKSVLNTFARGRTFLMTMSY 391
|||||:::|||||  |||||
orf133ng    DAGNDAATQRYSSFDPKDKDEDVTGNADKTLNCGRYGGTSKSVLNTFARGRTFLMTMSY 920

orf133.pep  KF 393
            ||
orf133ng    KF 922

```

The complete length ORF133ng nucleotide sequence <SEQ ID 881> is predicted to encode a

protein having amino acid sequence <SEQ ID 882>:

```

1  MRSSFLKPI  CFYLMGVMLY  HHSYAFDAGR  AGSERQIQVL  EDHVHKARRV
51  PKDKKVPTDA  RAVSTRQDVF  KSGENLDNIV  RSTPGAFTQQ  DKSSGIVSLN
101  IRGDSGFGRV  NTMVDGITQT  FYSTSTDAGR  AGSSSQFGAS  VDSNFIAGLD
151  VVKGSGFSGA  GINSIAGSAN  LRTLGVDDVV  QGNNTYGLLL  KGLTGTNSTK
201  GNAMAAIGAR  KWLESGASVG  VLYGHSRRGV  AQNYRVGGGG  QHIGNFGEY
251  LERRKQQYFV  QEGGLKFNAG  SGKWERDLQR  QWKTWKYKK  YEDPQLQKY
301  IEEHDKSWRE  NLAPOYDTP  IDPSGLKQOS  AGNLLMLEYD  GVFNKYTAQF
351  RDLNTRIGSR  KTIINRYQFN  YGLSLNPTYN  LNLTAAYNSG  RQKYPKGAQF
401  TGWGLLKDFE  TYNNAKILD  NNTATFRLPR  ETELQTTLGF  NYFHNEYGKN
451  RFPEELGLFF  DGPDPQDNGLY  SYLGRFKGKD  GLLPQKSTIV  QPAGSQFYNT
501  FYFDAALKKD  IYRLNYSINA  INYRFGEYIT  GYYSSENEFK  RAFGENSPAY
551  KEHCDPSCGL  YEPVLKKYGR  KRANNHVSIS  SADFGDYFMP  FAGYSRTHRM
601  PNQBMYYFSQ  IGDSGVHVAL  KPERANTWQF  GENTYKRGLL  KQDDILGLKL
651  VYGRSLDMY  JENYVYKWWO  LMGDSIPSWG  STGLAYTIRH  RNFKDKVIKH
701  GFLELNDYD  GRFFTMLSYA  YQKSTQPTNF  SDASEPNNA  SKEDQLKQY
751  GLSRVSALPR  DYRLLEVGR  WLGNKLLTGG  AMRYPGKSIR  ATADERYIDG
801  TNGGNTSNVR  QLGKRSIKQT  ETLAGQPLIF  DFYAAVEPKK  NLI FRAEVKN
851  LFDRYRIDPL  DAGNDAATQR  YSSFDPKDK  DEDVTGNADK  TLNCGRYGGT
901  SKSVLNTFAR  GRTFLMTMSY  KF*

```

A variant was also identified, being encoded by the gonococcal DNA sequence <SEQ ID 883>:

```

1  ATGAGATCTT  CTTTCGGGTT  GAAGCGGATT  TGTTTTTCAT  TTATGGGTGT
51  TATGCTATAT  CATCATAGTT  ATGCCGAAGA  TGCAGGCGCG  CGCGGCACGC
101  AGGCGCAGAT  ACAGGTTTTG  GAAGATGTGC  ACCTCAAGCG  GAAGCGCGTA
151  CCGAAAGACA  AAAAAGTGTT  TACCGATGCG  CGTGCGGAT  CGACCCGTca
201  gATGATGTT  AAATCCGGCG  AAAACCTCGA  CAACATCGTA  CGCAGCATAC
251  COSSTGCGTT  TACACAGCAA  GATPAAAGCT  CGGCGATTGT  GTCTTTGAAT
301  ATTCCGCGCG  ACACGCGGTT  CGGCGCGGTC  AATACGATGC  TGACAGGCAI
351  CACGCAAGC  TTTTATTGGA  CTCTTACGGA  TCGCGGCGAT  CGACGCGGTT
401  CATCTCAATT  CGGTGCATCT  GTCGACAGCA  ATTTTATTGC  CGGACTGATG
451  GTCGTCGAAG  GCAGCTTCAG  CGGCTCGGCA  GGCATCAACA  GCCTTGCCGG
501  TTGCGCGAAT  CTGCGGACTT  TAGGCGTGGA  TGAAGCTCGTT  CAGGCGAATA
551  ATACCTACGG  CTGCTGTGTA  AAGGTCTGTA  CGGCGACCAA  TTCAACCAA
601  GGTAAATCGA  TGGCGCGGAT  AGGTGCGGCG  AATGGCTGG  AAGCGGGAGC
651  GTCGTGCGT  GTGCTTTACG  GGCAACAGAG  CGCGCGCGCT  GCGCAAAATT
701  ACCGCGTGGG  CGGCGCGGGG  CAGCACAATG  GAAATTTTGG  TGAAGAAAT
751  CTGGAACGCG  GCAAACAGCA  ATATTTTGTA  CAAGAGGGGT  GTTTGAAATT
801  CAATGCGGCG  AGCGGAAAAA  GGGAAOAGGA  TTTGCAAAAG  CAATACTGGA
851  AAACAAGAGT  GTATAAAAAA  TACGAAGACC  CCAAGAAGCT  GCAGAAATAC
901  ATCGAAGAGC  ATGATAAAGG  CTGGCGGGAA  AACCTTGC3C  CGCAATACGA
951  CTCACGCCCC  ATCGATCGGT  CGGCGCTBAA  GCGACGATCG  CGGCACATC
1001  TGTTTAATTT  GGAATACGAC  GCGATTTTCA  ATAAATACAC  CGCGCAATTT
1051  CGCATTTTAA  ACACAGCAAT  CGCGAGCGCG  AAATCATCA  ACCGCAATTA
1101  TCAATTCNA  TACGGTTTGT  CTTTGAACCC  GTATACCAAC  CTCAATCTGA
1151  CCGGACGCTA  CAATTCGGGC  AGGCAGAAAT  ATCGCAAGAG  GGCGAAGTTT
1201  ACAGGCTGGG  GGCTTTTAAA  AGATTTTGAA  ACCTACACAA  ACAGCGAAAT
1251  CCTGACGCTC  AACACACGCG  CACCTTCGCG  GCTGCCCCGC  GAAACCGAGT
1301  TGCACACCC  TTTGGGCTTC  AATTTATTTCC  ACAACGAATA  CGGCAAAATC
1351  CGCTTTCCTG  AAGAATTGGG  GCTGTTTTTC  GACGCTCTCT  ATCAGGACAA
1401  CGGCTTTAT  TCCTATTGCG  GCGCGTTTAA  GGGCGATAAA  GGGCTGTTGC
1451  CTCAAAATC  AACCATTTGC  CAACGCGCGC  GCAGGCAATA  TTTCAACACG
1501  TTTCTACTTG  ATCGCGCGCT  CAATAAAGAC  ATTTACCGCT  TAACTACAGC
1551  CACCAATGCA  ATCACTTACC  GTTTCGCGCG  CGAATATACG  GCGTATTACG
1601  GCTCGGAAA  CGAATTACAG  CGGCGATTCG  GAGAAATCTC  GCGGCAATC
1651  AAGGACGAT  GCGACCGGAG  CTGCGACTTC  TATGAAACCG  TATTGAAAAA
1701  ATACGCGCAA  AAGCGCGCCA  ACAACATTC  AGTCAGACTT  AGTCGCGACT
1751  TCGCGGATTA  TTTGATGCGC  TTCGCGGCT  ATTCGCGGAC  ACACCGTATG

```

10

15

20

25

30

35

40

45

60

65

	orfl33-1	: : :: GNFGAEYLERKKQRYFVQEGALKFNDSGKWERDLQRQQWKYPYKNYNN-QELQKYIEE 220 230 240 250 260
5	orfl33ng-1.pep	310 320 330 340 350 360 HDKSWRENLAQYDITPIDPSGLKQASAGNLFKLEYDGVFNKYTAQFRDLNTRIGSKRII
10	orfl33-1	270 280 290 300 310 320 HDKSWRENLAQYDITPIDPSGLKQASAGNLFKLEYDGVFNKYTAQFRDLNTRIGSKRII
	orfl33ng-1.pep	370 380 390 400 410 420 NRNYQFNYGLSLNPFYTNLNLTAAYNSGRQKYPKGAKFTGWGLLKDFETYNNAKILDNLNT
15	orfl33-1	330 340 350 360 370 380 NRNYQFNYGLSLNPFYTNLNLTAAYNSGRQKYPKGAKFTGWGLLKDFETYNNAKILDNLNT
	orfl33ng-1.pep	430 440 450 460 470 480 ATFRLPRETQLTTLGFNYFHNHYGKRNFFPEELGLFFDGPDDNGLYSYSLGRFKGDKGLL
20	orfl33-1	390 400 410 420 430 440 ATFRLPRETQLTTLGFNYFHNHYGKRNFFPEELGLFFDGPDDNGLYSYSLGRFKGDKGLL
	orfl33ng-1.pep	490 500 510 520 530 540 PQKSTIVQAGSQYFNTFYFDAALKDIYRLNYSNTNAINYRFGGEYTYGYGSENEFKRAF
25	orfl33-1	450 460 470 480 490 500 PQKSTIVQAGSQYFNTFYFDAALKDIYRLNYSNTNAINYRFGGEYTYGYGSENEFKRAF
	orfl33ng-1.pep	550 560 570 580 590 600 GENSPAYKEHCDPSCGLYEPVLKKYKGRKRNHNSVSISADFGDYFMPFAGYSRTHMPNI
30	orfl33-1	510 520 530 540 550 560 GENSPAYKEHCDPSCGLYEPVLKKYKGRKRNHNSVSISADFGDYFMPFAGYSRTHMPNI
	orfl33ng-1.pep	610 620 630 640 650 660 QEMYFSQIGDSGVHTALKPERANTWQFGFNTYKGLLKQDDTLGLKLVGYRSRIDNYIHN
35	orfl33-1	570 580 590 600 610 620 QEMYFSQIGDSGVHTALKPERANTWQFGFNTYKGLLKQDDTLGLKLVGYRSRIDNYIHN
	orfl33ng-1.pep	670 680 690 700 710 720 VYKQWDLNGDIPSWVSGTGLAYTIRHRNFKDKVHKHGFLELNYDYGRTTNLSYAYQK
40	orfl33-1	630 640 650 660 670 680 VYKQWDLNGDIPSWVSGTGLAYTIRHRNFKDKVHKHGFLELNYDYGRTTNLSYAYQK
	orfl33ng-1.pep	730 740 750 760 770 780 STQPTNFSDASESFNNAKEDQLKQGYGLSRVSALPRDYGRLEVGTWRWLNKLTGGAMR
45	orfl33-1	690 700 710 720 730 740 STQPTNFSDASESFNNAKEDQLKQGYGLSRVSALPRDYGRLEVGTWRWLNKLTGGAMR
	orfl33ng-1.pep	790 800 810 820 830 840 YFGKSIRATAERYIDGTNGGNTSNFRQLGKRSIKQTTETLARQPLIFDYFAAYEPKKNLI
50	orfl33-1	750 760 770 780 790 800 YFGKSIRATAERYIDGTNGGNTSNFRQLGKRSIKQTTETLARQPLIFDYFAAYEPKKNLI
	orfl33ng-1.pep	850 860 870 880 890 900 FRAEVKNLFDRIYIDPLDAGNDAATQRYSSDFPKKDEVDVTNADKTLONGYGGTSKS
55	orfl33-1	810 820 830 840 850 860 FRAEVKNLFDRIYIDPLDAGNDAATQRYSSDFPKKDEVDVTNADKTLONGYGGTSKS
	orfl33ng-1.pep	910 920 VLTNFAARGTFLMTMSYKFX
60	orfl33-1	870 880 VLTNFAARGTFLMTMSYKFX

70 In addition, ORF133ng-1 is homologous to a TonB-dependent receptor in *H.influenzae*:

sp|P45114|YC17_HAEN PROBABLE TONB-DEPENDENT RECEPTOR H1217 PRECURSOR
>gi|1075372|p|I|G64110 transferrin binding protein 1 precursor (tbpl) homolog -
Haemophilus influenzae (strain Rd KW20) >gi|1574147 (U32801) transferrin binding
protein 1 precursor (tbpl) [Haemophilus influenzae] Length = 913
Score = 930 bits (2377), Expect = 0.0
Identities = 476/921 (51%), Positives = 619/921 (66%), Gaps = 72/921 (7%)

5 Query: 38 QVLEDDVHVKAARVFKDKVFTDARAVSTRQDVFKSGENLDIVRSIPGAFTQDKSSGIV 97
+ L + V R + DKK FT+A:A STR++VEK +D ++R3IPGAFTQDK SG+V
10 Sbjct: 29 ETLQGDIVVEKVI3NDKKPFTEAKAKSTRNVFKETOTIDQVIRSIPGAFTQDKSGSVV 98

Query: 98 SLNIRGD3GFGVRNMTVDGIGTITFYSTSDAGRAGSSQFASVDSNFIAGLDVVKGSFS 157
S+NIRG++G GRVNTMTVDG+TOTFYST+ D+G++GGSSQFGA++D NFIAG+DV K +FS
Sbjct: 89 SVNIRGNGGLGRVNTMTVDGVTOTFYSTALD9GSGSSQFGAIDFNFIAGVDVKNKSNFS 148

15 Query: 158 GSAGINSLAGSANLRTLGVDVQVXXXXXXXXXXXXXXXXXXXXXAMAAIGARKWLESGA 217
G++GIN+LAGSAN RTLG+DV+ M RKWL+G
Sbjct: 149 GASGINALAGSANFRTLGVDVITDDKPGIILKGMTGSNATKSNFMTMAAGRKWLNDGG 208

20 Query: 218 SVGVLYGHSRGRVAGVQVYRVGGGQHIGFEEYLERRKQOYFVQGGGLKFNAGSGKWERD 277
VGV++G+S+R V+Q+YR+ GGG+ + G++ L + K+ YF + G N G+W D
Sbjct: 209 YGVGVYGY3QREV3QDYRI- GGGERLASLGQDILAKEEAYF-RNAGYILNP-EGGWTPD 265

25 Query: 278 LQRQYWK-----TKWY-----KKYEDPOBLOK---YEE 303
L+++W +Y KK +D ++LQK IEE
Sbjct: 266 LSKKHWSCKNPDYQKNGDCSYRIGSAAKTRREILQELLTNGKKPKDIEKLQKNGDIE 325

Query: 304 HKSWSRENLAFOYDITPIDPSGLKQOSAGNLFKLEYDGVFNKYTAOFRDLNTRIGSRKII 363
DKS+ N QY + PI+P L+ +S +L K EY AQ R L+ +GSRKI
30 Sbjct: 326 TDKSFERN-KDOYSVAPIEPGSLQSRSRSHLLKFEYGDHQLGAOLRLTNKIGSRKIE 384

Query: 364 NRNYQFNYGLSLNPTNLNLTAAYNSGRQKYPKGAFTGWGLKDFETYNNAKILDLNNT 423
NRNYQ NY + N Y +LNL AA+N G+ YEK G FW + T N A I++NN+
Sbjct: 385 NRNYQVNYFNNNYSLDLNLMAAHNIGKTIYKGGFFAGWQVADKLITKNVANIVDINN3 444

35 Query: 424 ATRFLPRETELQTLTGNYFNHNEYGNRFEELGLFFDGDQDNGLYSY--LGRFKGDKG 481
TF LP+E +L+TTLGNYFY NEY KNRFEEL LF++ D GLYS+ GR+ G K
Sbjct: 445 HTFLPKEDILKTTLGNFYFTNEYSKNRFEEELSLFYNDASHDQGLYSHSKGRYSGTKS 504

40 Query: 482 LLPKSTIVQAGSQYNTFYFOALKKDIYRLNYSNTNAINRYFGGEYTGYYGSENEFKR 541
LLPQ+S I+QP+G Q F T YFD AL K IY LNY N +Y F GEY GY
Sbjct: 505 LLPORSVILQPSGKQKFTVYFOTALSKGIYHLYNSVNTFYAFNGEYVGY----- 555

45 Query: 542 AFGENSPAYKEHCDPSCGLYEPVLKKYKGRANNHVSISADFGDYFMPFAGYSRTHRMP 601
EN+ + +EP+L K G K+A NHS ++SA+ DYFMPF YSRTHRMP
Sbjct: 556 ---ENTAGQO-----INEPILKHSGHKKAENHSATLSAELSDYFMPFFYSRTHRMP 604

Query: 602 NIQEMFYSQIGDSGVHTALKPERANTWOPGENTYKKGKLDQDILGLKLVGYRSIDNYI 661
NIQEM+FSQ++GV+TALKPE++T+Q GENTYKKGK QD+LG+KLGVYRS I NYI
50 Sbjct: 605 NIQEMFYSQVSNAGVNTALKPEQSDTYQLGENTYKKGKLTQDDVLGVKLGVYRSFIKNIYI 664

Query: 662 HNYGVKWDLNDGIPSVWGSTGLAYTIRHNFKDKVKGHGFELNLDYDGRFFNLSYAY 721
HNYGV WW +P+W S G YTI H+N+K V K G ELE+NYD GRFF N+SYAY
Sbjct: 665 HNYGVVW--RDGMPTWASNGFKYTIHQNKYKEIVKKSQVLEINYDMGRFFANVSAY 722

55 Query: 722 QKSTOPTNFSASESPNNASKEDOLKQGYLSRVLSALPRDYGRLVGTGRVGNKILTLGGA 781
Q++ OPTN++DAS PNNAS+ED LKQGYLSRVLS LF+DYGRLV+GTRW KITLG A
Sbjct: 723 QRTNQTPTNADASPRPNNASQEDILKQGYLSRVLSMLFKDYGRLVGTGRVFNKILTLG 782

60 Query: 782 MRYFGKSIRATAERYIDGTNGNTSNVRQLGKRSIKQOTETIARQPLIFDFYAYEPKKN 841
RY+GKS RAT EE YI+G+ +R+ ++K+TE + +QP+I D + +YEP K+
Sbjct: 783 ARYFGKSKRATIEEYINSR-FKKNLTARENNYAVKKTEDIKKQPIILDLHVSYPEIKD 841

65 Query: 842 LIFRAEVKNLFDRRYIDPLDAGNDAATQRYSSFDKDKDEVTNCADKTLCKNGYGTGS 901
LI +AEV-NL D-RY-DPLDAGNDA+ORYYSS + +C D +C GG+
Sbjct: 842 LIIKAEVQNLDKRYVDPLDAGNDAASQRYSSLI-----NNSIECAQDSSAC---GGSD 892

Query: 902 KSVLTNFGARTRFLMTMSYKF 922
K+VL NFGARTR+++++YKF
70 Sbjct: 893 KTVLYNFGARTRIYLSNLYKF 913

The underlined motif in the gonococcal protein (also present in the meningococcal protein) is predicted to be an ATP/GTP-binding site motif A (P-loop), and the analysis suggests that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

5 Example 104

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 885>

```

1 ATGAACCTGA TTTACGTTA CATCATCGT CAAATGGCGG TTATGGCGGT
51 TTACGCGCTC CTTGCCTTCG TCGCTTTGTA CAGCTTTTTT GAAATCCTGT
101 ACGAAACCGG CAACCTCGGC AAAGGACAGT ACGGCATATG GGAATGCTG
151 GGCTACACCG CCTCAAAAT GCGCGCCGCG GCTACGAAAC TGATTCCTCT
201 CGCGCTCCTT ATCGCGCGAC TGCTCTCCCT CAGCGAGCTT GCGCGCGCA
251 GCGAAGTACG CGTCATCAAA GCGACGGGCA TGAGGACCAA AAAGCTGCTG
301 TTGATTCTGT CGCAGTTCGG TTTTATTTT GCTATTGCCA CGGTCGCGCT
351 CGGCGAATGG GTTGGGCGCA CACTGAGCCA AAAAGCCGAA AACATCAAG
401 CGCGCGCAT CAACGCAAA ATCAGCACCG GCAATACCGG CTTTGGCTG
451 AAGAAAAAAA ACAGCGTGAT CAATGTGCGC GAAATGTTCG CCGACCAT..

```

This corresponds to the amino acid sequence <SEQ ID 886; ORF112>:

```

1 MNLISRYIIR QMAVMAYVAL LAFLLALYSFF EILYETGNLG KSGYGIWEML
51 GYTALKMPAR AYELIPLAVL IGGVLSLSQL AAGSELTVIK ASGMSTKRL
101 LILSGFGFIF AIATVALGEW VAPTLSQKAE NIKAAAINKG ISTNTGLWL
151 KEKNSVINVR EMLFDH...

```

Further work revealed further partial nucleotide sequence <SEQ ID 887>:

```

1 ATGAACCTGA TTTACGTTA CATCATCGT CAAATGGCGG TTATGGCGGT
51 TTACGCGCTC CTTGCCTTCG TCGCTTTGTA CAGCTTTTTT GAAATCCTGT
101 ACGAAACCGG CAACCTCGGC AAAGGACAGT ACGGCATATG GGAATGCTG
151 GGCTACACCG CCTCAAAAT GCGCGCCGCG GCTACGAAAC TGATTCCTCT
201 CGCGCTCCTT ATCGCGCGAC TGCTCTCCCT CAGCGAGCTT GCGCGCGCA
251 GCGAAGTACG CGTCATCAAA GCGACGGGCA TGAGGACCAA AAAGCTGCTG
301 TTGATTCTGT CGCAGTTCGG TTTTATTTT GCTATTGCCA CGGTCGCGCT
351 CGGCGAATGG GTTGGGCGCA CACTGAGCCA AAAAGCCGAA AACATCAAG
401 CGCGCGCAT CAACGCAAA ATCAGCACCG GCAATACCGG CTTTGGCTG
451 AAGAAAAAAA ACAGCTTKAT CAATGTGCGC GAAATGTTCG CCGACCATAC
501 GCTTTTGGGC ATCAAAATTG GCGCGCGCAA CGATAAAAC GAATTGGCAG
551 AGGCAAGTGA AGCGAATTC GCGGTTTGA ACAGCGACGG CAGTGGCAG
601 TTGAAAAACA TCGCGCGCAG CAGCGTTGGC GAAGACAAAG TCGAGGCTC
651 TATTGCGGCT GAAGAAAACT GCGCGATTTC CGTCAACGCG AACCTGATGG
701 ACGTATTGCT CGTCAAAACC GACCAATGT CCGTGGGCGA ACTGACCAAC
751 TACATCGCGC ACCTCCAAAA CAACAGGCCA AACACCCGAA TCTACGCCAT
801 CGCATGGTGG CGCAAAATGG TTTACCCCGG CGCAGCGCTG GTGATGGCGG
851 TCGTGCCTTT TGCGCTTACC CCGCAACACA CCGCGCACGG CAATATGGCG
901 TTAAGACTCT TCGCGCGCAT CTGTSTCGGA TTGCTGTTCG AACCTGGCGG
951 ACGGCTCTTT GGGTTACCA GCCAATCGG...

```

This corresponds to the amino acid sequence <SEQ ID 888; ORF112-1>:

```

1 MNLISRYIIR QMAVMAYVAL LAFLLALYSFF EILYETGNLG KSGYGIWEML
51 GYTALKMPAR AYELIPLAVL IGGVLSLSQL AAGSELTVIK ASGMSTKRL
101 LILSGFGFIF AIATVALGEW VAPTLSQKAE NIKAAAINKG ISTNTGLWL
151 KEKNSXINVR EMLFDHTLLG IKIWARNDKN ELAEAEVADS AVLNSDGSWQ
201 LKNIRSTLGE EDKVEVSIAS EENWPISVKR NLMVDLVLPQ DQMSVGLTTP
251 YIRHLQNNQS NTRIYAIAWW RKLVIYPAAW VMLVAFAET PQTRRHGMNG
301 LKLGIGICKG LLFLHAGRLF GFTSL...

```

Computer analysis of this amino acid sequence predicts two transmembrane domains and gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF112 shows 96.4% identity over a 166aa overlap with an ORF (ORF112a) from strain A of *N. meningitidis*:

		10	20	30	40	50	60
5	orf112.pep	MNLI	SRYYIRQMAVM	VYALLAFL	ALYSFF	EILYETGNL	KGSYGIWEM
	orf112a	MNLI	SRYYIRQMAVM	VYALLAFL	ALYSFF	EILYETGNL	KGSYGIWEM
		10	20	30	40	50	60
10	orf112.pep	AYEL	IPLAVLIGGLV	LSQLAAGSE	LTVIKASGM	STKRL	LLLSQFGFI
	orf112a	AYEL	IPLAVLIGGLV	LSQLAAGSE	LTVIKASGM	STKRL	LLLSQFGFI
		70	80	90	100	110	120
15	orf112.pep	VAPT	LSQKAENIKAA	INGKISTGNT	GLWKEKNS	VINREML	PDH
	orf112a	VAPT	LSQKAENIKAA	INGKISTGNT	GLWKEKNS	VINREML	PDH
		130	140	150	160	170	180
20	orf112a	ELAE	AVEADSAVLN	SDGSQWLK	NIIRRTSLG	EDKVEVSI	AAEEXWFI
	orf112a	ELAE	AVEADSAVLN	SDGSQWLK	NIIRRTSLG	EDKVEVSI	AAEEXWFI
		190	200	210	220	230	240

The ORF112a nucleotide sequence <SEQ ID 889> is:

25	1	ATGAACCTGA	TTTCAOGTTA	CATCATCOGT	CAATATGCGG	TTATGCGGGT
	51	TTACGGCGCTC	CTTGCCCTTCC	TGCGTTTGTGA	CAGCTTTTTT	GAATCCTGT
	101	ACGAAACCGG	CACCTTCGGC	AAAGCGAGTT	ACGGCATATG	GGAAATGNTG
	151	GGNTACACCG	CCCTCAAAAT	GGCCGCGCGC	GCTACGAC	GGATGGCCCT
30	201	CGCGCTCTT	ATCGGGCGCA	TGCTCTCTT	CAGCCAGCT	GCOCGCGCA
	251	GGCAACTGAN	CGCTCATCAA	CGCAGCGGCA	TGAGCACCA	AAAGCTGCTG
	301	TTGATTCTGT	CGCAGTTCGG	TTTTATTTTT	GCTATTGCCA	CCGTGCGGCT
	351	CGCGCGAATGG	GTTGCGGCCA	CAGTGAAGCA	AAAGCGCGAA	AACATCAAG
	401	CGCGCGCAT	CAACGGCRAA	ATCAGTACCG	GCAATACCGG	CCTTTGGCTG
	451	AAAGAAAAAA	ACAGCATTAT	CAATGTGCGC	GAATGTGTC	CGACCATAC
35	501	CTGCTGGGC	ATTAATAATCT	GGGCGCGCAA	GATATAAACC	GAACCTGGCAG
	551	AGGCAGTGG	AGCGGATTC	GCGGTTTTGA	ACAGCGACGG	CAGTTGGCAG
	601	TTGAAACAA	TCCGCGCAG	CACGCTTGGC	GAAGACAAG	TCGAGGCTC
	651	TATTGCGGCT	GAAGAAANT	GGCGGATTTC	CGTCAACGC	AACCTGATGG
	701	ACGTATTGCT	CGTCAACCC	GACCAATGT	CCGTGCGGCA	ACTGACACCC
40	751	TACATCGCGC	ACCTCCAAAN	NNACGCCAA	ACACCCGAA	TCTACGCCAT
	801	CGCATGGTG	CGCAATTTG	TTTACCGCG	CGACGCTGG	GGATGATGG
	851	TGTCGCGCT	TGCTTTAGC	CGCGAACA	CCCGCACGG	CAATATGGCG
	901	TTAANANTCT	TCGGGGCAT	CTGTCTCGGA	TGCTGTCTCC	ACCTTCGCGG
	951	NCGGCTCTTC	NGGTTTACCA	GCCAACCTCA	CGGCATCCG	CCCTTCCTCG
45	1001	NCGGCGCACT	ACCTACCATA	GCTTTCGCT	TGCTGCGCGT	TTGCTGATA
	1051	CGCAACAGG	AAAAACGCTA	A		

This encodes a protein having the amino acid sequence <SEQ ID 890>:

	1	MNLI	SRYYIRQMAVM	VYALLAFL	ALYSFF	EILYETGNL	KGSYGIWEM
	51	GYTALKMXAR	AYELMPLAVL	IGGLVXSQ	L	AAGSELXV	IKASGMSTKRL
50	101	LILSQFGFI	AIATVALGEW	VAPTLSQKAE	NIKAAEANG	ISTGNTGLWL	
	151	KEKNSINVR	EMLPDHTLLG	IKIWARNDKN	ELAEAVEADS	AVLNSDGSQW	
	201	LKNIRRTSLG	EDKVEVSI	AAEEXWFI	PISVVKR	NLMDVLVLPK	DQMSVGLPFT
	251	YIRHLQXXSQ	NTRYALIAWW	RKLVPYPAAM	VMALVAFAT	PQTRHGMNG	
	301	LKFGGICLG	LLFHLA	GRRLF	XFTSQLYGIF	PFLXGLPFI	AFALLAVWLI
55	351	RKQER					

ORF112a and ORF112-1 show 96.3% identity in 326 aa overlap:

	orf112a.pep	MNLI	SRYYIRQMAVM	VYALLAFL	ALYSFF	EILYETGNL	KGSYGIWEM	MXGYTALKMXAR
	orf112-1	MNLI	SRYYIRQMAVM	VYALLAFL	ALYSFF	EILYETGNL	KGSYGIWEM	IGLYTALKMPAR
60	orf112a.pep	AYEL	MPLAVLIGGLV	XSQLAAGSE	LXV	IKASGMSTKRL	LLLSQFGFI	AIATVALGEW

ORF112 shows 95.8% identity over 166aa overlap with a predicted ORF (ORF112ng) from *N.*

25	orf112.pep	MNLSIRSYIRQMVAWAVALLAFALYSFFFEILLYETGNLKGSGYIWEMLGYTAKMPAR	60
	orf112.ng	MNLSIRSYIRQMVAWAVALLAFALYSFFFEILLYETGNLKGSGYIWEMLGYTAKMPAR	60
30	orf112.pep	AYELIPLAVLIGGLVSLSQLAAGSELTVIKASGMSKTKLLLLISQFGFIFAIATVALGEW	120
	orf112.ng	AYELMPLAVLIGGLSLSQLAAGSELVLIKASGMSKTKLLLLISQFGFIFAIARAALGEW	120
	orf112.pep	VAPTLISQKAENIKAAINGKISTGNTGVLWLEKKSVINVRMLPDH	166
	orf112.ng	VAPTLISQKAENIKAAINGKISTGNTGVLWLEKTSIINVRGMLPDHTLLGIGKIWARNDKN	180

35	1	ATGAACCTGTA	TTTACAGTTA	CATCATCTGGC	CAAAATGGCGG	TTATGGCGGT
	51	TTTACGGCGGTC	CTTGGCTTCTC	TTCGATTCGGA	CGATCTTTTT	GAAATCTGTG
	101	ACGAAACACCG	CAACCTCGGC	AAAGCGATGT	ACGGCATATG	GGAATAGCTG
	151	GGCTACACCG	CCCTCAAAAT	CGGCGCGCGT	GCCTACGAAC	TCATGCCCTC
	201	CGCGCTGCTC	ATCGCGCGAC	TGGCTCTCTT	CAGCGAGCTT	CGCGCGCGCA
40	251	CGGAACCTGC	CGTCATCAAA	GCGACGCGCT	TAGCGACCAA	AAAGCTGCTG
	301	TGATATTCTG	CTCAGTTCGG	TTTATTTT	GCAATTGGCG	CCGTGCGCCT
	351	CGCGGATGAG	GTTCGGGCCA	CGCTGAGCCA	CGGACCGGAA	AACATCAAGg
	401	CGCGCGCGGA	ATACAGCGG	ATACAGCGG	CTTTTCCTCT	CTTTTCCTCT
	451	AAAGAAATAA	CAGCGATTAT	CAATCGGCGG	CGAATTCTGC	CGGACCATAC
	501	CGTTTGTGGC	ATACAAATTT	GGGCGCGCAA	CGATAAAACG	GAATTGGGAT
	551	AGGCAGTGGA	AGGCCGATTCC	CGCGTTTGGT	ACAGCAGACGG	CAGCTGGGAC
	601	TTGAAAAACA	TCGCGCGCAG	CATCATTTGA	CACAGACAAA	TCGAACATC
	651	cgCGCGCGCG	GAGAAACCTT	gCGCATATTG	CGTCAAGCAG	AACCTGATGG
	701	ACGTATTGCT	CGTCAAGCCG	GACCAATATG	CGTGCGGCGA	GCTGACCAAC
50	751	TACATCCGCG	CGCTACAAA	ACACAGCCAA	ACACGCCAAA	TCTAGCCAT
	801	CGCATGGTGG	AGTAACTGCT	TTTACCGCGT	ACGCGCATAG	GTCATGGCGC
	851	TGTTTGGCTT	CGCTTTTACG	CGGCAAAACA	CGGCGCACAG	CATATGGGCT
	901	TTAAGCTGTA	GGGTTTGGTA	CTGTCTCGGA	CGGCGGCTCA	CGGCGGCTCA
	951	CAGGCTTCTC	GGGTTTCCA	CGCAATCTCA	CGGACACCCA	CCCTTCTGGC
55	1001	CGCGGCGACT	CGCTACCAAT	CGCTTGCCTC	TGCTCGCTGT	TTGGCTGATA
	1051	CGCAACRAGG	AAAAACGTTG	b.		

1	MNLSRYIR	QMANVAYAL	LAFLALYSFF	EILYETGLNV	KGSYGIWML
51	GYALKRMPAR	AVELMPLAVL	TGGLSLASLQ	ARGSELALV	ASGMRKRL
101	LILSOFQIF	AAVALAGVGE	VAPTLISQRAE	NKIAAINGK	LSTNGTLKL
151	KEKTSINVR	KMLPDHLLI	KLWARNDRN	MLDAVLAVS	AVLSDGSGWQ
201	LKNRRISNG	TOKIETSAAA	EETWPIAVRR	NLMOLVLLPK	QMSYVELTT
251	YRHLQNNQS	NTQYIAIWM	RKTLVYPAWM	VNALVAFAT	POTTRHNGE
301	LKI.EGGTGIS	LILHILAGRLF	GGTSYVPEGL	PEFLAGLPTI	AFALLAWML

351 RKQEK*

ORF112ng and ORF112-1 show 94.2% identity in 326 aa overlap:

		10	20	30	40	50	60
5	orf112ng	MNLSIRYII	QMAVMAYV	ALLAFALY	SFFELI	YETGNL	GKGSYGIW
	orf112-1	MNLSIRYII	QMAVMAYV	ALLAFALY	SFFELI	YETGNL	GKGSYGIW
		10	20	30	40	50	60
10	orf112ng	70	80	90	100	110	120
	orf112-1	AYELIPLAV	LIGGLASL	SQAAGSEL	TVIKASGM	STKLLIL	LSQFGFPA
		70	80	90	100	110	120
15	orf112ng	130	140	150	160	170	180
	orf112-1	VAPTLSQKA	ENIKAAAI	NGKISTGT	GLWKEKTS	IINVRGML	PDHTLLGI
		130	140	150	160	170	180
20	orf112ng	190	200	210	220	230	240
	orf112-1	ELAEAEAD	SAVLNSD	GSWQLKN	IRRSIMGT	DKIETSA	AREETWPI
		190	200	210	220	230	240
25	orf112ng	250	260	270	280	290	300
	orf112-1	DQMSVGEL	TTYIRHL	QNNQNT	QIYAI	AWWRKLV	YPVAAW
		250	260	270	280	290	300
30	orf112ng	310	320	330	340	350	
	orf112-1	LKLPGGIC	LGLFHL	LAGRLP	GPPTSQ	LYGTPPF	LALAGLPT
		310	320				

This analysis suggests that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their
 40 epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

It will be appreciated that the invention has been described by means of example only, and that
 modifications may be made whilst remaining within the spirit and scope of the invention.

TABLE I – PCR primers

ORF	Primer	Sequence	Restriction sites
ORF 1	Forward	CGCGGATCCGCTAGC-GGACACACTTATTTCGG	BamHI-NheI XhoI
	Reverse	CCCCTCGAG-CCAGCGGTAGCCTAATT	
ORF 2	Forward	GCGGATCCCATATG-TTTGATTTCGGTTTGGG	BamHI-NdeI XhoI
	Reverse	CCCCTCGAG-GACGGCATAACGGCG	
ORF 2-1	Forward	GCGGATCCCATATG-TTTGATTTCGGTTTGGG	BamHI-NdeI XhoI
	Reverse	CCCCTCGAG-TGATTACGGACGCGCA	
ORF 4	Forward	GCGGATCCCATATG-TGCGGAGGTCAAAAAGAC	BamHI-NdeI XhoI
	Reverse	CCCCTCGAG-TTTGGCTGCGCCTTC	
ORF 5	Forward	GGAATTCCATATGGCCATGG-TGGAAGGCGCACAAAC	NdeI-NcoI BamHI XhoI
	Reverse	CGGGATCC-ATGGAAGGCGCACAAAC CCCCTCGAG-GACTGTGCAAAAACGG	
ORF 6	Forward	CGCGGATCCCATATG-ACCCGTCAATCTCTGCA	BamHI-NdeI XhoI
	Reverse	CCCCTCGAG-TGCGCCGAACACTTTC	
ORF 7	Forward	CGCGGATCCGCTAGC-GCGCTGCTTTTTGTTC	BamHI-NheI XhoI
	Reverse	CCCCTCGAG-TTTCAAAATATATTTGCGGA	
ORF 8	Forward	GCGGATCCCATATG-GCTCAACTGCTTCGTAC	BamHI-NdeI XhoI
	Reverse	CCCCTCGAG-AGCAGGCTTTGGCGC	
ORF 9	Forward	CGCGGATCCCATATG-CCGAAGGAAGTCGGAAA	BamHI-NdeI XhoI
	Reverse	CCCCTCGAG-TTTCGAGGTTTTCGGG	
ORF 10	Forward	GCGGATCCCATATG-GACACAAAAGAAATCCTC	BamHI-NdeI XhoI
	Reverse	CCCCTCGAG- TAATGGGAACCTTGTTTT	
ORF 11	Forward	GCGGATCCCATATG-GCGGTCAACCTCTACG	BamHI-NdeI XhoI
	Reverse	CCCCTCGAG-GGAAACGACTTCGCC	
ORF 13	Forward	CGCGGATCCCATATG-GCTCTGCTTTCCGCGC	BamHI-NdeI XhoI
	Reverse	CCCCTCGAG-AGGCTGTGTGATAATAAG	
ORF 15	Forward	GGAATTCCATATGGCCATGG-GCGGACACTGACAG	NdeI-NcoI BamHI XhoI
	Reverse	CGGGATCC-TGCGGGACACTGACAGG CCCCTCGAG-AGGTTGGCCTTGCTATG	
ORF 17	Forward	GGAATTCCATATGGCCATGG -TTGCCGGCCTGTTCTG	NdeI-NcoI

	Forward Reverse	CGGGATCC-ATTGCCGGCCTGTTGC CCCGCTCGAG-AAGCAGTTGTACAGC	BamHI XhoI
ORF 18	Forward Reverse	GCGGATCCCATATG-ATTTTGCTGCATTTGGAT CCCGCTCGAG-TCTTCCAATTTCTGAAAGC	BamHI-NdeI XhoI
ORF 19	Forward Forward Reverse	GGAATTCATATGGCCATGG -TCGCCAGTGTTTTACC CGGGATCC-TTCGCCAGTGTTTTACC CCCGCTCGAG-GGTGTTTTGAAGCTGCC	NdeI-NcoI BamHI XhoI
ORF 20	Forward Forward Reverse	GGAATTCATATGGCCATGG -TCGCCGCGGGTATG CGGGATCC-TTCGCCGCGGGTATG CCCGCTCGAG-CGGCGAGCGAGAGCA	NdeI-NcoI BamHI XhoI
ORF 22	Forward Forward Reverse	GGAATTCATATGGCCATGG-TGATTAAATCAAAAAGGTCT CGGGATCC-ATGATTAAATCAAAAAGGTCTAAACC CCCGCTCGAG-ATTATGATAGCGGCC	NdeI-NcoI BamHI XhoI
ORF 23	Forward Reverse	CGCGGATCCCATATG-GATGTTTCTGTTTCAGAC CCCGCTCGAG-TTTAAACCGATAGTTAAACG	BamHI-NdeI XhoI
ORF 24	Forward Forward Reverse	GGAATTCATATGGCCATGG -TGATGCCGGAATGGTG CGGGATCC-ATGATGCCGGAATGGTG CCCGCTCGAG-TGTCAGCGTGCGCA	NdeI-NcoI BamHI XhoI
ORF 25	Forward Reverse	GCGGATCCCATATG-TATCGCAAACTGATTGC CCCGCTCGAG-ATCGATGGAATAGCCG	BamHI-NdeI XhoI
ORF 26	Forward Reverse	GCGGATCCCATATG -CAGCTGATCGACTATTG CCCGCTCGAG-GACATCGGCGCGTTTT	BamHI-NdeI XhoI
ORF 27	Forward Forward Reverse	GGAATTCATATGGCCATGG-AGACCTATTCTGTTTA CGGGATCC- CAGACCTATTCTGTTTATTTAATC CCCGCTCGAG-GGGTTCGATTAAATAACCAT	NdeI-NcoI BamHI XhoI
ORF 28	Forward Forward Reverse	GGAATTCATATGGCCATGG-ACGGCTGTACGTTGATGT CGGGATCC-AACGGCTGTACGTTGATG CCCGCTCGAG-TTTGTACAGGAATTCGCG	NdeI-NcoI BamHI XhoI
ORF 29	Forward Forward Reverse	GCGGATCCCATATG -AACGGTTTGGATGCCCG CGCGGATCCGCTAGC-AACGGTTTGGATGCCCG CCCGCTCGAG-TTTGTCTAAGTTCTGATATG	BamHI-NdeI BamHI-NheI XhoI
ORF 32	Forward Reverse	CGCGGATCCCATATG-AATACTCCTCCTTTTG CCCGCTCGAG-GCGTATTTTGTGATGCTTTG	BamHI-NdeI XhoI
ORF 33	Forward Reverse	GCGGATCCCATATG -ATTGATAGGGATCGTATG CCCGCTCGAG-TTGATCTTTCAAACGGCC	BamHI-NdeI XhoI

ORF 35	Forward	GCGGATCCCATATG-TTCAGAGCTCAGCTT	BamHI-NdeI BamHI-NheI XhoI
	Forward	CGCGGATCCGCTAGC-TTCAGAGCTCAGCTT	
	Reverse	CCCGCTCGAG-AAACAGCCATTGAGCGA	
ORF 37	Forward	GCGGATCCCATATG-GATGACGTATCGGATTTT	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-ATAGCCCGCTTTCAGG	
ORF 58	Forward	CGCGGATCCGCTAGC-TCCGAACGCGAGTGGAT	BamHI-NheI XhoI
	Reverse	CCCGCTCGAG-AGCATTGTCCAAGGGGAC	
ORF 65	Forward	GGAATTCCATATGGCCATGG -TGCTGTATCTGAATCAAG	NdeI-NcoI BamHI XhoI
	Forward	CGGGATCC-TTGCTGTATCTGAATCAAGG	
	Reverse	CCCGCTCGAG-CCGCATCGGCAGACA	
ORF 66	Forward	GCGGATCCCATATG-TACGCATTACCGCGG	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-TGGATTTTGACAGATGG	
ORF 72	Forward	CGCGGATCCCATATG- AATGCAGTAAAAATATCTGA	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-GCCTGAGACCTTTGCAA	
ORF 73	Forward	GCGGATCCCATATG-AGATTTTTCGGTATCGG	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-TTCATCTTTTCATGTTGG	
ORF 75	Forward	GCGGATCCCATATG- TCTGTCTTTCAAACGGC	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-TTTGTTTTTGCAAGACAG	
ORF 76	Forward	GATCAGCTAGCCATATG-AAACAGAAAAAACCGC	NheI-NdeI BamHI
	Reverse	CGGGATCC-TTACGGTTTGACACCGTT	
ORF 79	Forward	CGCGGATCCCATATG-GTTTCGCGCGCGG	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-GTGCTGATGCGCTTCG	
ORF 83	Forward	GCGGATCCCATATG-AAAACCTTGCTGCTGC	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-GCCGCCCTTGGGGC	
ORF 84	Forward	GCGGATCCCATATG-GCAGAGATCTGTTTG	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-GTTTGCGATCCGACCA	
ORF 85	Forward	CGCGGATCCCATATG- GCGGTTTGGGGCGGA	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-TCGGCGCGCGGGC	
ORF 89	Forward	GGAATTCCATATGGCCATGG- CCATACCTTCTTATCA	NdeI-NcoI BamHI XhoI
	Forward	CGGGATCC-GCCATACCTTCTTATCAGAG	
	Reverse	CCCGCTCGAG-TTTTGTGCGATTAGAAAAAGC	
ORF 97	Forward	GCGGATCCCATATG-CATCTGCCAGCGAAC	BamHI-NdeI

	Reverse	CCCGCTCGAG-TTCGCCTACGGTTTTTTC	XhoI
ORF 98	Forward	GCGGATCCCATATG-ACGGTAACTGCGG	BamHI-NdeI
	Reverse	CCCGCTCGAG-TTGTTGTCGGGCAAATC	XhoI
ORF 100	Forward	GCGGATCCCATATG-TCGGGCATTACACCG	BamHI-NdeI
	Reverse	CCCGCTCGAG-ACGGGTTTCGGCGGAA	XhoI
ORF 101	Forward	GCGGATCCCATATG-ATTTATCAAGAAACCTC	BamHI-NdeI
	Reverse	CCCGCTCGAG-TTTTCCGCCCTTTCATGT	XhoI
ORF 102	Forward	GCGGATCCCATATG-GCAGGGCTGTTTTACC	BamHI-NdeI
	Reverse	CCCGCTCGAG-AAACGGTTTGAAACACGAC	XhoI
ORF 103	Forward	GCGGATCCCATATG-AACCACGACATCAC	BamHI-NdeI
	Reverse	CCCGCTCGAG-CAGCCACAGGACGGC	XhoI
ORF 104	Forward	GCGGATCCCATATG-ACGTGGGGAACGC	BamHI-NdeI
	Reverse	CCCGCTCGAG-GCGGCGTTTGACCGGC	XhoI
ORF 105	Forward	GCGGATCCCATATG-ACCAAATTTCAAACCCCTC	BamHI-NdeI
	Reverse	CCCGCTCGAG-TAAACGAATGCCGTCCAG	XhoI
ORF 106	Forward	GCGGATCCCATATG-AGGATAACCGACGGCG	BamHI-NdeI
	Reverse	CCCGCTCGAG-TTGTTCGCCGATGATGT	XhoI
ORF 109	Forward	GCGGATCCCATATG-GAAGATTTATATATAACTCG	BamHI-NdeI
	Reverse	CCCGCTCGAG-ATCAGCTTCGAACCGAAG	XhoI
ORF110	Forward	AAAGAATTC-ATGAGTAAATCCCGTAGATCTCCC	EcoRI
	Reverse	AAACTGCAG-GGAAAACCATCCGCACTCTGCC	PstI
ORF111	Forward	AAAGAATTC-GCACCGCAAAAGGCAAAAACCGCA	EcoRI
	Reverse	AAACTGCAG-TCTGCGCGTTTTCGGCAGGGTGG	PstI
ORF113	Forward	AAAGAATTC-ATGAACAAAACCCCTATCGTGTGATTTTCAACCG	EcoRI
	Reverse	AAACTGCAG-TTACGAATGCTGCTTGTCTGACCGTACTG	PstI
ORF115	Forward	AAAGAATTC-TTGCTTGTGCAACAGAAAAAGACGG	EcoRI
	Reverse	AAAAAGTCGAC-CTATTTTTAGGGGCTTTGCTGTGTTGAAAAGCCTGCC	SalI
ORF119	Forward	AAAGAATTC-TACAACATGTATCAGGAAAACCAATACCG	EcoRI
	Reverse	AAACTGCAG-TTATGAAAACAGGCGAGGGCGGTTTTGCG	PstI
ORF120	Forward	AAAGAATTC-GCAAGGCTACCCCAATCCGCGGTG	EcoRI
	Reverse	AAACTGCAG-CGGTTTGGCTGCTGGCCGCTTGAT	PstI
ORF121	Forward	AAAGAATTC-GCCTTGGTCTGGCTGGTTTTGCG	EcoRI
	Reverse	AAACTGCAG-TCATCCGCCACCCACCTCGGCCATCCATC	PstI

ORF122	Forward Reverse	AAAAAGTCGAC-ATGTCCTACCGCAAGCAGTTCCTC AA <u>ACTG</u> CAG-TCAGGAACACAAACGATGACGAATATCCGTATC	Sall PstI
ORF125	Forward Reverse	AAAGAATTC-GCGCTGTTTTTTGCGGCGGCGTAT AA <u>ACTG</u> CAG-CGCGGTTTCAAGCGAAAAAGTCG	EcoRI PstI
ORF126	Forward Reverse	AAAGAATTC-GCGGAAACGGTCAAG AA <u>ACTG</u> CAG-TTAATCTTGCTTCCGATATAC	EcoRI PstI
ORF127	Forward Reverse	AAAGAATTC-ATGACTGATAATCGGGGTTTACG AAAAAGTCGAC-CTTAAGTAACTGCAGTCCTTATC	EcoRI Sall
ORF128	Forward Reverse	AAAGAATTC-ATGCAAGCTGTCGCTACAGGCC AA <u>ACTG</u> CAG-CTATTGCAATGCGCGCCCGGGAATGTTGAGAGGGC	EcoRI PstI
ORF129	Forward Reverse	AAAGAATTC-ATGGATTTTCGTTTGACATTATTACGAATACCG AA <u>ACTG</u> CAG-TTATTTTTTGATGAAATTTTGGGCGG	EcoRI PstI
ORF130	Forward Reverse	AAAGAATTC-GCAGTACTTGCCATTCCTCGGTGCG AA <u>ACTG</u> CAG-CTCCGGATCGTCTGTAAACGCATT	EcoRI PstI
ORF 131	Forward Reverse	GCGGATCCCATATG-GAAATTCGGGCAATAAAAT CCCGCTCGAG-CCAGCGGACGCGTTC	BamHI-NdeI XhoI
ORF 132	Forward Reverse	GCGGATCCCATATG-AAAGAAGCGGGTTTG CCCGCTCGAG-CCAATCTGCCAGCCGT	BamHI-NdeI XhoI
ORF 133	Forward Reverse	CGCGGATCCCATATG-GAAGATGCAGGCGCGG CCCGCTCGAG-AACTTGTAGCTCATCGT	BamHI-NdeI XhoI
ORF 134	Forward Reverse	GCGGATCCCATATG-TCTGTGCAAGCAGTATTG CCCGCTCGAG-ATCCTGTGCCAATGCG	BamHI-NdeI XhoI
ORF 135	Forward Reverse	GCGGATCCCATATG-CGCTCTGAAAAAGCTTT CCCGCTCGAG-AAATACCGCTGAGGATG	BamHI-NdeI XhoI
ORF 136	Forward Reverse	CGCGGATCCCGTAGC-ATGAAGCGGCGTATAGCC CCCGCTCGAG-TTCGAATATTGGAACTTTT	BamHI-NheI XhoI
ORF 137	Forward Reverse	CGCGGATCCCATATG-GGCACGGCGGGAATA CCCGCTCGAG-ATAACGGTATGCCGCC	BamHI-NdeI XhoI
ORF 138	Forward Reverse	GCGGATCCCATATG-TTTCGTTTACAATTCAGGC CCCGCTCGAG-CGGCGTTTATAGCGG	BamHI-NdeI XhoI
ORF 139	Forward Reverse	GCGGATCCCATATG-GCTTTTTTGCGGTAATG CCCGCTCGAG-TAACGTTTCCGTGCGTTT	BamHI-NdeI XhoI

ORF 140	Forward Reverse	GCGGATCCCATATG-TTGCCACAGGCAGC CCCGCTCGAG-GACGATGGCAACAGC	BamHI-NdeI XhoI
ORF 141	Forward Reverse	GCGGATCCCATATG-CCGTCTGAAGCAGTCT CCCGCTCGAG-ATCTGTTGTTTTAAATATT	BamHI-NdeI XhoI
ORF 142	Forward Reverse	GCGGATCCCATATG-GATAATTCTGGTAGTGAAG CCCGCTCGAG-AAACGTATAGCCTACCT	BamHI-NdeI XhoI
ORF 143	Forward Reverse	GCGGATCCCATATG-GATACCGCTTTGAACCT CCCGCTCGAG-AATGGCTTCCGCAATATG	BamHI-NdeI XhoI
ORF 144	Forward Reverse	GCGGATCCCATATG-ACCTTTTACAACGTTTGC CCCGCTCGAG-AGATTGTTGTTGTTTTTCG	BamHI-NdeI XhoI
ORF 147	Forward Reverse	GCGGATCCCATATG-TCTGTCTTTCAAACGGC CCCGCTCGAG-TTTGTTTTGCAAGACAG	BamHI-NdeI XhoI

NB:

- restriction sites are underlined
- for ORFs 110-130, where the ORF itself carries an *EcoRI* site (eg. ORF122), a *SalI* site was used in the forward primer instead. Similarly, where the ORF carries a *PstI* site (eg. ORFs 115 and 127), a *SalI* site was used in the reverse primer.

TABLE II – Summary of cloning, expression and purification

ORF	PCR/cloning	His-fusion expression	GST-fusion expression	Purification
orf 1	+	+	+	His-fusion
orf 2	+	+	+	GST-fusion
orf 2.1	+	n.d.	+	GST-fusion
orf 4	+	+	+	His-fusion
orf 5	+	n.d.	+	GST-fusion
orf 6	+	+	+	GST-fusion
orf 7	+	+	+	GST-fusion
orf 8	+	n.d.	n.d.	
orf 9	+	+	+	GST-fusion
orf 10	+	n.d.	n.d.	
orf 11	+	n.d.	n.d.	
orf 13	+	n.d.	+	GST-fusion
orf 15	+	+	+	GST-fusion
orf 17	+	n.d.	n.d.	
orf 18	+	n.d.	n.d.	
orf 19	+	n.d.	n.d.	
orf 20	+	n.d.	n.d.	
orf 22	+	+	+	GST-fusion
orf 23	+	+	+	His-fusion
orf 24	+	n.d.	n.d.	
orf 25	+	+	+	His-fusion
orf 26	+	n.d.	n.d.	
orf 27	+	+	+	GST-fusion
orf 28	+	+	+	GST-fusion
orf 29	+	n.d.	n.d.	
orf 32	+	+	+	His-fusion
orf 33	+	n.d.	n.d.	
orf 35	+	n.d.	n.d.	
orf 37	+	+	+	GST-fusion
orf 58	+	n.d.	n.d.	
orf 65	+	n.d.	n.d.	
orf 66	+	n.d.	n.d.	
orf 72	+	+	n.d.	His-fusion
orf 73	+	n.d.	+	n.d.
orf 75	+	n.d.	n.d.	
orf 76	+	+	n.d.	His-fusion
orf 79	+	+	n.d.	His-fusion
orf 83	+	n.d.	+	n.d.
orf 84	+	n.d.	n.d.	

orf 85	+	n.d.	+	GST-fusion
orf 89	+	n.d.	+	GST-fusion
orf 97	+	+	+	GST-fusion
orf 98	+	n.d.	n.d.	
orf 100	+	n.d.	n.d.	
orf 101	+	n.d.	n.d.	
orf 102	+	n.d.	n.d.	
orf 103	+	n.d.	n.d.	
orf 104	+	n.d.	n.d.	
orf 105	+	n.d.	n.d.	
orf 106	+	+	+	His-fusion
orf 109	+	n.d.	n.d.	
orf 110	+	n.d.	n.d.	
orf 111	+	+	n.d.	His-fusion
orf 113	+	+	n.d.	His-fusion
orf 115	n.d.	n.d.	n.d.	
orf 119	+	+	n.d.	His-fusion
orf 120	+	+	n.d.	His-fusion
orf 121	+	n.d.	n.d.	
orf 122	+	+	n.d.	His-fusion
orf 125	+	+	n.d.	His-fusion
orf 126	+	+	n.d.	His-fusion
orf 127	+	+	n.d.	His-fusion
orf 128	+	n.d.	n.d.	
orf 129	+	+	n.d.	His-fusion
orf 130	+	n.d.	n.d.	
orf 131	+	+	+	n.d.
orf 132	+	+	+	His-fusion
orf 133	+	n.d.	+	GST-fusion
orf 134	+	n.d.	n.d.	
orf 135	+	n.d.	n.d.	
orf 136	+	n.d.	n.d.	
orf 137	+	n.d.	+	GST-fusion
orf 138	+	n.d.	+	GST-fusion
orf 139	+	n.d.	n.d.	
orf 140	+	n.d.	n.d.	
orf 141	+	n.d.	n.d.	
orf 142	+	n.d.	n.d.	
orf 143	+	n.d.	n.d.	
orf 144	+	n.d.	+	n.d.
orf 147	+	n.d.	n.d.	

CLAIMS

1. A protein comprising an amino acid sequence selected from the group consisting of SEQ IDs 2, 4, 6, and 8.
2. A nucleic acid molecule which encodes a protein according to claim 1.
3. A nucleic acid molecule according to claim 2, comprising a nucleotide sequence selected from the group consisting of SEQ IDs 1, 3, 5, and 7.
4. A protein comprising an amino acid sequence selected from the group consisting of SEQ IDs 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 228, 230, 232, 234, 236, 238, 240, 242, 244, 246, 248, 250, 252, 254, 256, 258, 260, 262, 264, 266, 268, 270, 272, 274, 276, 278, 280, 282, 284, 286, 288, 290, 292, 294, 296, 298, 300, 302, 304, 306, 308, 310, 312, 314, 316, 318, 320, 322, 324, 326, 328, 330, 332, 334, 336, 338, 340, 342, 344, 346, 348, 350, 352, 354, 356, 358, 360, 362, 364, 366, 368, 370, 372, 374, 376, 378, 380, 382, 384, 386, 388, 390, 392, 394, 396, 398, 400, 402, 404, 406, 408, 410, 412, 414, 416, 418, 420, 422, 424, 426, 428, 430, 432, 434, 436, 438, 440, 442, 444, 446, 448, 450, 452, 454, 456, 458, 460, 462, 464, 466, 468, 470, 472, 474, 476, 478, 480, 482, 484, 486, 488, 490, 492, 494, 496, 498, 500, 502, 504, 506, 508, 510, 512, 514, 516, 518, 520, 522, 524, 526, 528, 530, 532, 534, 536, 538, 540, 542, 544, 546, 548, 550, 552, 554, 556, 558, 560, 562, 564, 566, 568, 570, 572, 574, 576, 578, 580, 582, 584, 586, 588, 590, 592, 594, 596, 598, 600, 602, 604, 606, 608, 610, 612, 614, 616, 618, 620, 622, 624, 626, 628, 630, 632, 634, 636, 638, 640, 642, 644, 646, 648, 650, 652, 654, 656, 658, 660, 662, 664, 666, 668, 670, 672, 674, 676, 678, 680, 682, 684, 686, 688, 690, 692, 694, 696, 698, 700, 702, 704, 706, 708, 710, 712, 714, 716, 718, 720, 722, 724, 726, 728, 730, 732, 734, 736, 738, 740, 742, 744, 746, 748, 750, 752, 754, 756, 758, 760, 762, 764, 766, 768, 770, 772, 774, 776, 778, 780, 782, 784, 786, 788, 790, 792, 794, 796, 798, 800, 802, 804, 806, 808, 810, 812, 814, 816, 818, 820, 822, 824, 826, 828, 830, 832, 834, 836, 838, 840, 842, 844, 846, 848, 850, 852, 854, 856, 858, 860, 862, 864, 866, 868, 870, 872, 874, 876, 878, 880, 882, 884, 886, 888, 890, & 892..
5. A protein having 50% or greater sequence identity to a protein according to claim 4.

6. A protein comprising a fragment of an amino acid sequence selected from the group consisting of SEQ IDs 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 228, 230, 232, 234, 236, 238, 240, 242, 244, 246, 248, 250, 252, 254, 256, 258, 260, 262, 264, 266, 268, 270, 272, 274, 276, 278, 280, 282, 284, 286, 288, 290, 292, 294, 296, 298, 300, 302, 304, 306, 308, 310, 312, 314, 316, 318, 320, 322, 324, 326, 328, 330, 332, 334, 336, 338, 340, 342, 344, 346, 348, 350, 352, 354, 356, 358, 360, 362, 364, 366, 368, 370, 372, 374, 376, 378, 380, 382, 384, 386, 388, 390, 392, 394, 396, 398, 400, 402, 404, 406, 408, 410, 412, 414, 416, 418, 420, 422, 424, 426, 428, 430, 432, 434, 436, 438, 440, 442, 444, 446, 448, 450, 452, 454, 456, 458, 460, 462, 464, 466, 468, 470, 472, 474, 476, 478, 480, 482, 484, 486, 488, 490, 492, 494, 496, 498, 500, 502, 504, 506, 508, 510, 512, 514, 516, 518, 520, 522, 524, 526, 528, 530, 532, 534, 536, 538, 540, 542, 544, 546, 548, 550, 552, 554, 556, 558, 560, 562, 564, 566, 568, 570, 572, 574, 576, 578, 580, 582, 584, 586, 588, 590, 592, 594, 596, 598, 600, 602, 604, 606, 608, 610, 612, 614, 616, 618, 620, 622, 624, 626, 628, 630, 632, 634, 636, 638, 640, 642, 644, 646, 648, 650, 652, 654, 656, 658, 660, 662, 664, 666, 668, 670, 672, 674, 676, 678, 680, 682, 684, 686, 688, 690, 692, 694, 696, 698, 700, 702, 704, 706, 708, 710, 712, 714, 716, 718, 720, 722, 724, 726, 728, 730, 732, 734, 736, 738, 740, 742, 744, 746, 748, 750, 752, 754, 756, 758, 760, 762, 764, 766, 768, 770, 772, 774, 776, 778, 780, 782, 784, 786, 788, 790, 792, 794, 796, 798, 800, 802, 804, 806, 808, 810, 812, 814, 816, 818, 820, 822, 824, 826, 828, 830, 832, 834, 836, 838, 840, 842, 844, 846, 848, 850, 852, 854, 856, 858, 860, 862, 864, 866, 868, 870, 872, 874, 876, 878, 880, 882, 884, 886, 888, 890, & 892..
7. An antibody which binds to a protein according to any one of claims 4 to 6.
8. A nucleic acid molecule which encodes a protein according to any one of claims 4 to 6.
9. A nucleic acid molecule according to claim 8, comprising a nucleotide sequence selected from the group consisting of SEQ IDs 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133, 135, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 227, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, 253, 255, 257, 259, 261, 263, 265, 267, 269, 271, 273, 275, 277, 279, 281, 283, 285, 287, 289,

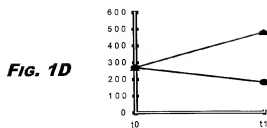
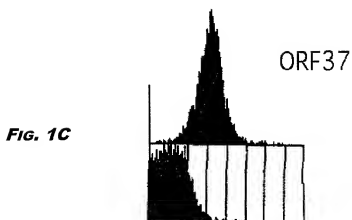
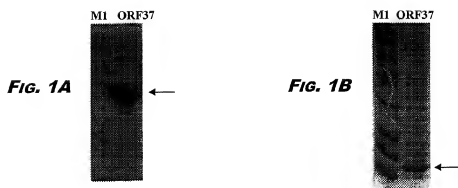
291, 293, 295, 297, 299, 301, 303, 305, 307, 309, 311, 313, 315, 317, 319, 321, 323, 325, 327, 329,
331, 333, 335, 337, 339, 341, 343, 345, 347, 349, 351, 353, 355, 357, 359, 361, 363, 365, 367, 369,
371, 373, 375, 377, 379, 381, 383, 385, 387, 389, 391, 393, 395, 397, 399, 401, 403, 405, 407, 409,
411, 413, 415, 417, 419, 421, 423, 425, 427, 429, 431, 433, 435, 437, 439, 441, 443, 445, 447, 449,
5 451, 453, 455, 457, 459, 461, 463, 465, 467, 469, 471, 473, 475, 477, 479, 481, 483, 485, 487, 489,
491, 493, 495, 497, 499, 501, 503, 505, 507, 509, 511, 513, 515, 517, 519, 521, 523, 525, 527, 529,
531, 533, 535, 537, 539, 541, 543, 545, 547, 549, 551, 553, 555, 557, 559, 561, 563, 565, 567, 569,
571, 573, 575, 577, 579, 581, 583, 585, 587, 589, 591, 593, 595, 597, 599, 601, 603, 605, 607, 609,
611, 613, 615, 617, 619, 621, 623, 625, 627, 629, 631, 633, 635, 637, 639, 641, 643, 645, 647, 649,
10 651, 653, 655, 657, 659, 661, 663, 665, 667, 669, 671, 673, 675, 677, 679, 681, 683, 685, 687, 689,
691, 693, 695, 697, 699, 701, 703, 705, 707, 709, 711, 713, 715, 717, 719, 721, 723, 725, 727, 729,
731, 733, 735, 737, 739, 741, 743, 745, 747, 749, 751, 753, 755, 757, 759, 761, 763, 765, 767, 769,
771, 773, 775, 777, 779, 781, 783, 785, 787, 789, 791, 793, 795, 797, 799, 801, 803, 805, 807, 809,
811, 813, 815, 817, 819, 821, 823, 825, 827, 829, 831, 833, 835, 837, 839, 841, 843, 845, 847, 849,
15 851, 853, 855, 857, 859, 861, 863, 865, 867, 869, 871, 873, 875, 877, 879, 881, 883, 885, 887, 889,
& 891..

10. A nucleic acid molecule comprising a fragment of a nucleotide sequence selected from the
group consisting of SEQ IDs 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39,
41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91,
20 93, 95, 97, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133,
135, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173,
175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, 211, 213,
215, 217, 219, 221, 223, 225, 227, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, 253,
255, 257, 259, 261, 263, 265, 267, 269, 271, 273, 275, 277, 279, 281, 283, 285, 287, 289, 291, 293,
25 295, 297, 299, 301, 303, 305, 307, 309, 311, 313, 315, 317, 319, 321, 323, 325, 327, 329, 331, 333,
335, 337, 339, 341, 343, 345, 347, 349, 351, 353, 355, 357, 359, 361, 363, 365, 367, 369, 371, 373,
375, 377, 379, 381, 383, 385, 387, 389, 391, 393, 395, 397, 399, 401, 403, 405, 407, 409, 411, 413,
415, 417, 419, 421, 423, 425, 427, 429, 431, 433, 435, 437, 439, 441, 443, 445, 447, 449, 451, 453,
455, 457, 459, 461, 463, 465, 467, 469, 471, 473, 475, 477, 479, 481, 483, 485, 487, 489, 491, 493,
30 495, 497, 499, 501, 503, 505, 507, 509, 511, 513, 515, 517, 519, 521, 523, 525, 527, 529, 531, 533,
535, 537, 539, 541, 543, 545, 547, 549, 551, 553, 555, 557, 559, 561, 563, 565, 567, 569, 571, 573,
575, 577, 579, 581, 583, 585, 587, 589, 591, 593, 595, 597, 599, 601, 603, 605, 607, 609, 611, 613,
615, 617, 619, 621, 623, 625, 627, 629, 631, 633, 635, 637, 639, 641, 643, 645, 647, 649, 651, 653,
655, 657, 659, 661, 663, 665, 667, 669, 671, 673, 675, 677, 679, 681, 683, 685, 687, 689, 691, 693,
35 695, 697, 699, 701, 703, 705, 707, 709, 711, 713, 715, 717, 719, 721, 723, 725, 727, 729, 731, 733,

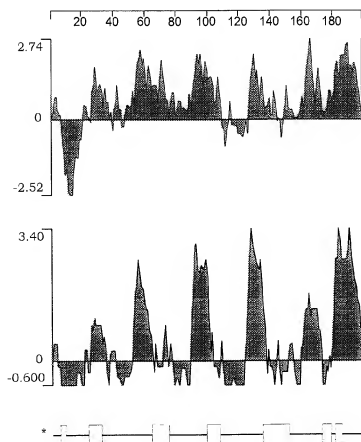
735, 737, 739, 741, 743, 745, 747, 749, 751, 753, 755, 757, 759, 761, 763, 765, 767, 769, 771, 773, 775, 777, 779, 781, 783, 785, 787, 789, 791, 793, 795, 797, 799, 801, 803, 805, 807, 809, 811, 813, 815, 817, 819, 821, 823, 825, 827, 829, 831, 833, 835, 837, 839, 841, 843, 845, 847, 849, 851, 853, 855, 857, 859, 861, 863, 865, 867, 869, 871, 873, 875, 877, 879, 881, 883, 885, 887, 889, & 891..

- 5 11. A nucleic acid molecule comprising a nucleotide sequence complementary to a nucleic acid molecule according to any one of claims 8 to 10.
12. A nucleic acid molecule comprising a nucleotide sequences having 50% or greater sequence identity to a nucleic acid molecule according to any one of claims 8-11.
13. A nucleic acid molecule which can hybridise to a nucleic acid molecule according to any
10 one of claims 8-12 under high stringency conditions.
14. A composition comprising a protein, a nucleic acid molecule, or an antibody according to any preceding claim.
15. A composition according to claim 14 being a vaccine composition or a diagnostic composition.
- 15 16. A composition according to claim 14 or claim 15 for use as a pharmaceutical.
17. The use of a composition according to claim 14 in the manufacture of a medicament for the treatment or prevention of infection due to Neisserial bacteria.

1/24

FIGURE 1

2/24

Fig 1E

3/24

FIGURE 2

Fig. 2A

M1 ORF5

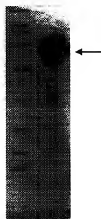
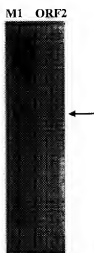
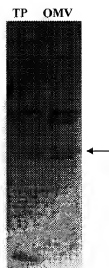
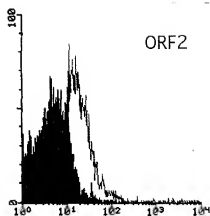


Fig. 2B

TP



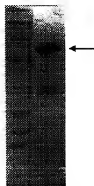
4/24

FIGURE 3**Fig. 3A****Fig. 3B****Fig. 3C****Fig. 3D**

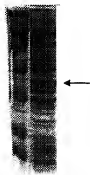
5/24

FIGURE 4**Fig. 4A**

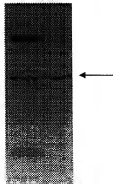
M1 ORF15

**Fig. 4B**

M2 ORF15

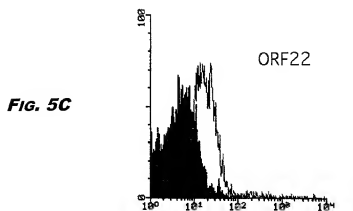
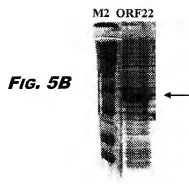
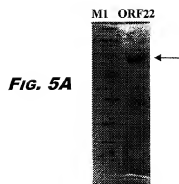
**Fig 4C**

TP OMV



6/24

FIGURE 5



7/24

FIGURE 6

Fig. 6A

M1 ORF28

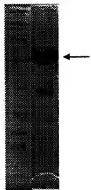


Fig. 6B

M2 ORF28

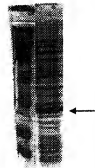


FIGURE 7

Fig. 7A

M1 ORF32

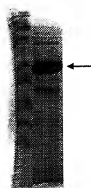
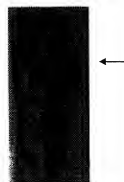
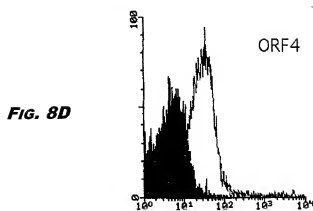
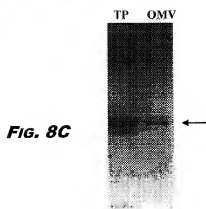
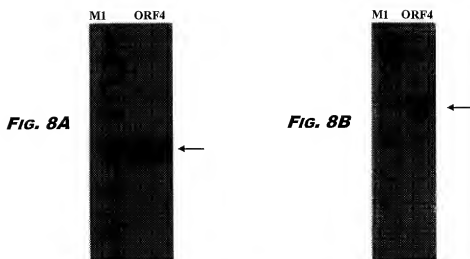


Fig. 7B

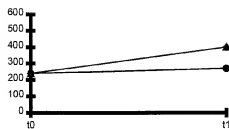
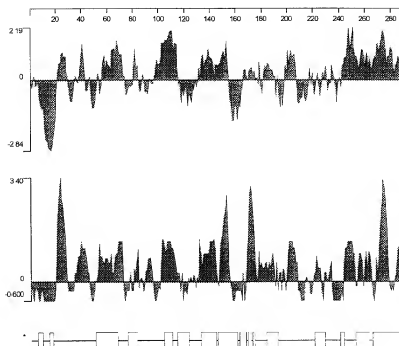
M1 ORF32



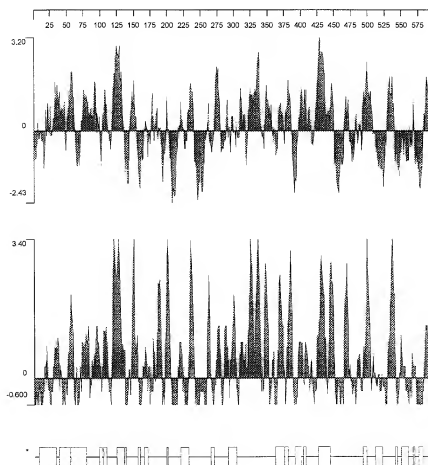
8/24

FIGURE 8

9/24

FIG. 8E**FIG. 8F**

10/24

FIGURE 9

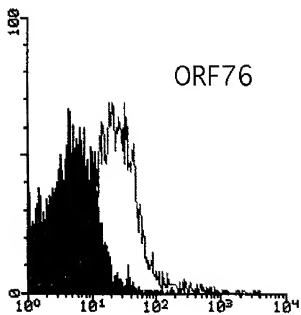
11/24

FIGURE 10**Fig. 10A**

M1 ORF76

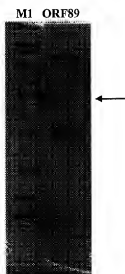
**Fig. 10B**

TP OMV

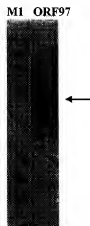
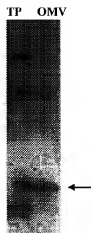
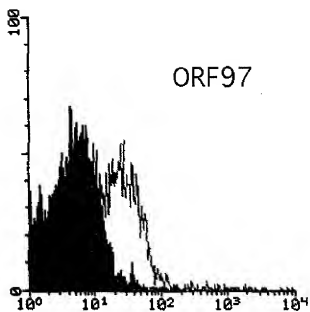
**Fig. 10C**

12/24

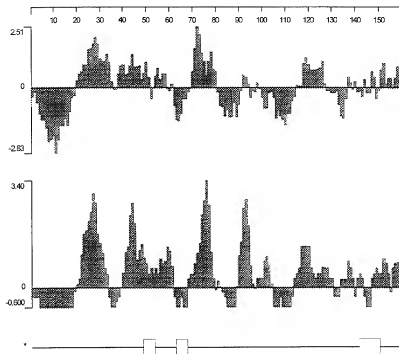
FIGURE 11



13/24

FIGURE 12**Fig. 12A****Fig. 12B****Fig. 12C****Fig. 12D**

14/24

Fig. 12E

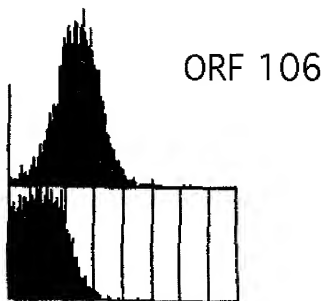
15/24

FIGURE 13**Fig. 13A**

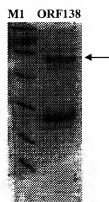
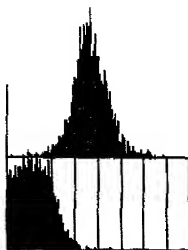
M1 ORF106

**Fig. 13B**

M2 ORF106

**Fig. 13C**

16/24

FIGURE 14***FIG. 14A******FIG. 14B***

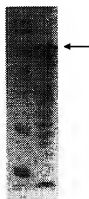
17/24

FIGURE 15***Fig. 15A***

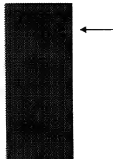
M1 ORF23

***Fig. 15B***

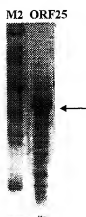
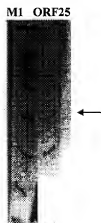
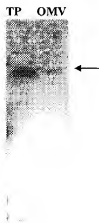
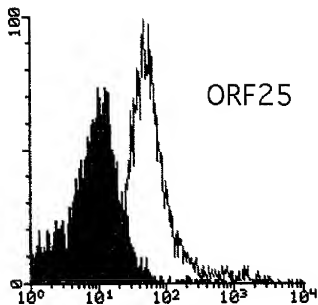
M2 ORF23

***Fig 15C***

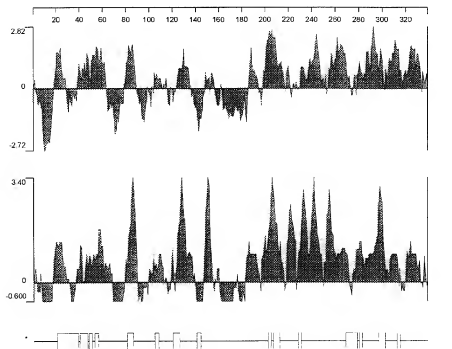
TP OMV



18/24

FIGURE 16**Fig. 16A****Fig. 16B****Fig. 16C****Fig. 16D**

19/24

FIG. 16E

20/24

FIGURE 17**Fig. 17A**

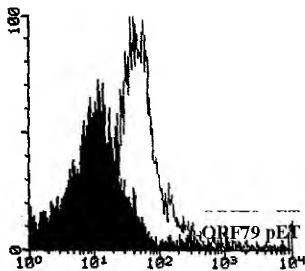
M1 ORF27

**Fig. 17B**

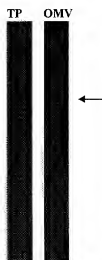
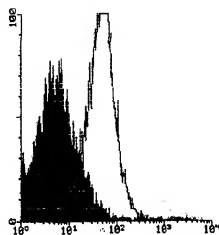
M2 ORF27



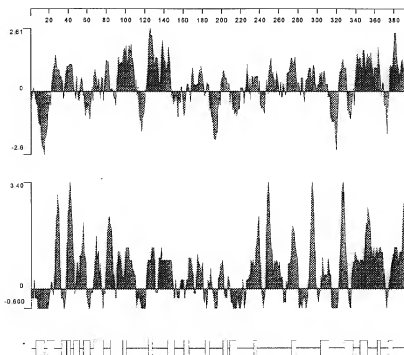
21/24

FIGURE 18**Fig. 18A****Fig. 18B**

22/24

FIGURE 19**Fig. 19A****Fig. 19B****Fig. 19C**

23/24

Fig 19D

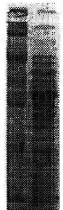
24/24

FIGURE 20**Fig. 20A**

M1 ORF132

**Fig. 20B**

M2 ORF132

**Fig. 20C**

ORF132